

AN ANALYSIS OF THE MORALITY OF INTENTION IN NUCLEAR DETERRENCE, WITH SPECIAL REFERENCE TO FINAL RETALIATION

JEFFREY ALOYSIUS ZINK

MERTON COLLEGE

HILARY 1990

A thesis submitted in partial fulfillment
of the requirements for the degree of D.Phil.
(62,000 words, inc. citations)

ABSTRACT

Jeffrey Aloysius Zink
Merton College

D.Phil.
Hilary 1990

AN ANALYSIS OF THE MORALITY OF INTENTION IN NUCLEAR DETERRENCE, WITH SPECIAL REFERENCE TO FINAL RETALIATION

Quite apart from its apparent political obsolescence, the policy of nuclear deterrence is vulnerable to attack for its seemingly obvious immorality. Nuclear war is blatantly immoral, and nuclear deterrence requires a genuine intention to resort to the nuclear retaliation which would precipitate such a war. Therefore, since it is wrong to intend that which is wrong to do, deterrence is immoral.

This thesis seeks to examine the nature of the deterrent intention as a means of verifying the soundness of the above deontological argument. This examination is carried out by first suggesting an acceptable notion of intention in general and then, after analysing the views of deterrent intention by other writers, proceeding to demonstrate the uniqueness of that intention. Having done this, and having explored the possibility that deterrence need not contain a genuine intention to retaliate, the thesis moves on to suggest and defend a moral principle which states that endeavours requiring the formation of an immoral intention may nevertheless be moral. Called the Principle of Double Intention (and based on the Principle of Double Effect), it offers a method for the moral assessment of agents who form immoral intentions within larger contexts. By applying this principle to nuclear deterrence, it is demonstrated that agents who undertake such a policy may be morally justified in doing so, provided certain conditions are met.

The thesis closes with a refutation of the objection that an agent cannot rationally form an intention (such as that required in deterrence) which he has no reason to carry out. By highlighting the objection's reliance on a claimed isomorphism between intention and belief, it is shown that the objection, while generally sound, does not apply to the special case of nuclear deterrence. The conclusion suggests a framework for disarmament which results in a deterrent force structure which is both strategically effective and morally acceptable.

To Diane, whose love and friendship
give meaning to my life.

PREFACE

Very late on one cold November night in North Dakota, during my six years as a bombardier flying B-52's and stationed on full alert readiness as the United States' first line of retaliatory deterrent defence, I was awakened by the blaring of the klaxon horns ordering us to scramble to the aircraft. While practice exercises were common, they were almost never scheduled while the bomber crews slept. So as we clamoured out of bed, into flight suits and out into the bitter cold to start the engines in preparation for imminent takeoff, the one thought which we never allowed ourselves to dwell on sprang up: this was no exercise. Nuclear war was about to begin.

As it happened, it was an exercise, although it was precipitated by an erroneous indication that the Soviet Union had launched its missiles. But the shock of that moment brought to life the genuine horror which would be the result of failed deterrence, a horror which seemed to transcend moral justification. But even accepting that conclusion, I still felt that deterrence itself was nonetheless justified, that what we were accomplishing out there on the frozen prairie, the prevention of enemy aggression, was morally acceptable. But I could not imagine a sound argument to support that claim.

As odd as it might seem, thinking about the problem of nuclear deterrence

was not part of the regime of the aircrews whose job it was to enforce that policy. We were concerned about delivery accuracy, fuel loads, times enroute to target, but never about the enormous consequences of utilising our training. Indeed, it was only after leaving the B-52 crew force that I had the chance to contemplate those consequences: I had the good fortune to be able to return to the study of moral philosophy, first as a student and then later as a teacher at the United States Air Force Academy. This change gave me the unique opportunity to develop a formal theoretical structure for thinking about the problem of nuclear deterrence. I have had the further good fortune to concentrate on that problem as a doctoral student at Merton College. This thesis is the culmination of my research into the problem, and represents what I hope is only the first phase of a continuing study into the issues of war, deterrence, and the morality of international affairs. The thesis is unique in that it contains the philosophical deliberations of one who has been, and will again be, working at the front lines of deterrence.

Completion of this research would literally have not been possible without the patient guidance of a number of scholars here at Oxford who have supervised my research, all of whom I gratefully acknowledge: John Finnis was overly kind to my initial and meager attempts at analysis, and indeed first suggested that I might focus on the problem of deterrent intention. Anthony Kenny generously gave me a term of his very limited time as I struggled through the arguments at the core of the thesis. Joseph Raz also made time for me, offering a critical look at the theoretical underpinnings of my ideas. Finally, and most of all, I must thank Jonathan Glover, who as my primary supervisor offered me the magical combination of intense critique and warm encouragement, the former forcing me to refine my hazy reasoning, the latter giving me the drive to continue.

For the extraordinary opportunity to read philosophy at Oxford I am deeply indebted to Colonel Malham Wakin, a brilliant professor, colleague and friend whose

truly unique gift for teaching has ignited and inspired literally thousands of students over the past thirty years, opening their eyes to the study of philosophy and showing them the crucial importance of ethics in the military profession.

Lastly, I wish to thank my parents, Joseph and Cecelia Zink, who encouraged and inspired all their children to excel at whatever they do. For that, and so much more, each of their children shall always be grateful.

Jeffrey A. Zink
Merton College

AN ANALYSIS OF THE MORALITY OF INTENTION
IN NUCLEAR DETERRENCE, WITH SPECIAL REFERENCE
TO FINAL RETALIATION

PART I: INTRODUCTORY REMARKS

1 The Moral Problem of Nuclear Deterrence 2

PART II: PRELIMINARY CONCEPTS

2 Some Definitions and Assumptions 13
3 A Plausible Notion of Intention in General 26

PART III: THE ROLE OF DETERRENT INTENTION

4 Can Deterrence Be Based on a Bluff? 49
5 Other Views of Deterrent Intention 67

PART IV: A REVISED NOTION OF DETERRENT INTENTION

6 The Unique Nature of Deterrent Intention 100
7 A Dualistic Model 109

PART V: THE PRINCIPLE OF DOUBLE INTENTION

8 The Principle of Double Effect 122
9 The Principle of Double Intention 145
10 Applying the Principle to Nuclear Deterrence 170
11 Agent Rationality and the Secondary Intention 182

PART VI: CONCLUDING REMARKS

12 Toward an Ideal Deterrent 207

Bibliography 214

Every man ought to endeavour peace, as far as he has hope of obtaining it.

--Thomas Hobbes

PART I: INTRODUCTORY REMARKS

1: THE MORAL PROBLEM OF NUCLEAR DETERRENCE

---

1. THE MORALITY OF WAR AND DETERRENCE 4

2. METHODS OF MORAL ANALYSIS 5

3. THE AIM OF THE THESIS 8

4. LARGER IMPLICATIONS 10

1: THE MORAL PROBLEM OF NUCLEAR DETERRENCE*

Hiroshima unrolled east to west in the cross hairs of Thomas Ferebee's Norden bombsight. The bomb-bay doors were open. The radio tone ended, the bomb dropped, Ferebee unclutched his sight. The arming wires pulled out to start Little Boy's clocks. The first combat atomic bomb fell away from the plane.

At 8:15 a.m. on Monday, the 6th of August 1945, as the bombardier Major Thomas Ferebee released the only weapon in the bomb-bay of the B-29, Enola Gay, he unleashed more than the destructive power of the newly developed bomb on the city of Hiroshima (although that alone would have been quite enough).[1] He simultaneously ushered in an era of radically altered international strategy, one in which the Clausewitzian view of war as an extension of politics by other means was replaced by the sober realisation that war was no longer a policy option for nuclear powers. Instead, emphasis shifted to achieving peace and security not by wielding military might after the fact, but by announcing beforehand deterrent threats to use atomic weapons

---

* *The opinions expressed in this thesis are those of the author. They are not necessarily the opinions of the USAF Institute of Technology, the United States Air Force, or the United States Government.*

[1] The narrative of the bombing is taken from Rhodes, p. 709.

if necessary. Thus the age of nuclear deterrence had begun. For the first two combat atomic bombs have also been the last.

1. THE MORALITY OF WAR AND DETERRENCE

Although issues of warfare have been the subject of moral debate since at least the time of St Augustine, the advent of nuclear power has added both a new dimension and a new urgency to that debate. Prior to 1945, warfare of the scale which could threaten the very existence of humanity itself was unthinkable. But the development and proliferation of weapons capable of swift and massive destruction has brought the unthinkable to the forefront of our collective consciousness. In a graphic introduction to the problem of nuclear weapons, Jonathan Schell writes that 'they are a pit into which the whole world can fall–a nemesis of all human intentions, actions, and hopes. Only life itself, which they threaten to swallow up, can give the measure of their significance.'[2]

The debate over nuclear weapons is not confined to the acceptability of their use, although that often is the starting point of the controversy since the potential for indiscriminate destruction disproportionate to the victory to be won naturally gives rise to the question of whether any use would satisfy the *jus in bello* criteria of Just-War Theory. For even if their use is prohibited, there remains a question of whether the possession of nuclear weapons may nevertheless be justified. And this of course opens the question of the morality of nuclear deterrence.

It is possible to examine the morality of deterrence itself from a number of different perspectives. One may, for example, approach the question from a consequentialist foundation, and argue that the benefits (or harms) of maintaining a

---

[2]Schell, p. 3.

deterrent posture outweigh other considerations, rendering deterrence an acceptable (or unacceptable) option.[3] Alternatively, one may undertake a deontological examination of deterrence, usually by analysing the intentions of the deterring agent.[4]

Examinations of the deontology of deterrence almost exclusively return a negative verdict on the policy, showing that forming the intention to retaliate--and thereby committing genocide on an unprecedented scale--is immoral. The present thesis is aimed at offering an altered deontological view of deterrent intention which will seek to demonstrate that maintaining the admittedly immoral intention to retaliate may be justified, not on the basis of the consequences of forming that intention, but because of the nature and purpose of the encompassing endeavour of deterrence which the agent undertakes.

2. METHODS OF MORAL ANALYSIS

The large question confronting those who think about the morality of strategic policy in a nuclear age is a most obvious one: Is deterrence by threat of nuclear retaliation a morally justifiable policy? It is a question which admits of but three answers, one positive, one negative, and the third a retort that the question itself is faulty. As a way of introducing the primary area of focus of this thesis, it may be of some interest to examine briefly the various routes by which one might arrive at one of the three answers.

To take the last answer first, the retort against the coherence of the question would most likely be offered by someone who holds a nihilistic view of morality, at

---

[3]Douglas Lackey (pp. 189-231) offers a consequentialist attack on deterrence, while Gregory Kavka (1978, pp. 285-302) suggests a consequentialist defence of the policy.

[4]For deontological views opposed to the acceptability of nuclear deterrence, one may choose from the works of (to name just a few) Germain Grisez (1982, pp. 9-24), Gerald Dworkin (pp. 445-60) and most recently Finnis et al. (esp. pp. 77-98). Perhaps the only deontology-based work supporting deterrence yet published is Kemp (1987a, pp. 276-94).

least with respect either to international relations or, more particularly, to war among nations.[5] That is, he would contend that issues of war or international diplomacy are amoral, and thus cannot be properly discussed from within the context of a moral framework. A particular policy may be judged right or wrong on the basis of its efficacy, say, or its coherence within a larger strategic framework, but it cannot be measured against any moral yardstick. One can no more ask about the morality of a particular foreign policy than one can ask about the shape of blue.

Despite the fact that many nations throughout history have apparently held such a view of morality, it seems highly implausible.[6] In any case, we may safely leave it outside the present realm of discussion.

The second, negative, answer to the question of the moral justifiability of nuclear deterrence is usually predicated on the claim that actual use of nuclear weapons is morally prohibited, i.e., that deterrence is wrong because the use of nuclear weapons, upon which deterrence is founded, is wrong.[7] This claim in turn is the result of one of two positions, either a general pacifist rejection of war in any form and therefore nuclear war,[8] or the view that nuclear weapons in particular are immoral because (for example) their use violates the Just-War Theory criteria of discrimination and proportionality.[9] This second and more focused criticism of deterrence comprises the majority of objections to the policy.

---

[5]Richard Wasserstrom (pp. 1636-43) provides a concise analysis and critique of this form of moral nihilism, arguing that it is fundamentally irrational.

[6]Examples here are by no means confined to blatantly evil empires such as Nazi Germany. In discussing both the United States' decision in 1949 to develop thermonuclear warheads and the 1962 Cuban missile crisis, former U.S. Secretary of State Dean Acheson frankly and unapologetically remarks that in each case 'moral talk did not bear on the problem.' Quoted in Wasserstrom, p. 1637.

[7]Although it is at least theoretically possible that one could hold that actual use is justifiable, but that nuclear deterrence itself is immoral, I, like Gerald Dworkin (p. 445), have found no one who actually supports such a stance.

[8]For a concise description of the various forms of pacifism, especially as they relate to nuclear war and deterrence, see Teichman, pp. 1-24 and 118-122.

[9]For a thorough and critical analysis of the criteria of Just-War Theory, see Childress, esp. pp. 434-39.

Finally, one may claim that nuclear deterrence is morally justifiable. Although the moralists who have supported this view have taken a number of different approaches to that conclusion, we may classify them into two broad categories, those who argue that the use of nuclear weapons, and thus deterrence, is justifiable, and those who admit that use is wrong but argue that the deterrent threat to use is nevertheless acceptable. In the first group are those, like David Gauthier, who argue that the retaliatory use of nuclear weapons is acceptable and therefore deterrence is permitted, and those, like David Fisher, who argue more modestly that since some use of nuclear weapons may be legitimate deterrence is justified.[10]

To put forth a pro-use defence of deterrence is to travel the moral rough road: The launch of nuclear weapons either in retaliation or as a first strike seems to be blatantly immoral given the expected genocide, especially if one accepts the escalation hypothesis that any retaliatory use of nuclear weapons will inevitably lead to full-scale exchange. As a result, most moralists who seek to defend deterrence fall into the second broad category, that is, they accept that use of nuclear weapons is wrong, but argue that nuclear deterrence is justified. The arguments employed within this category vary significantly, but may be loosely grouped under two headings, (1) those who claim that deterrence need not include the genuine intention to retaliate, and (2) those who claim that the intention, while necessary, is not immoral despite the fact that the act intended is wrong. Included in (1) are various bluffing theories of deterrence which we shall have occasion to examine in Chapter 4, while (2) includes arguments which call into question the applicability of the wrongful intentions principle, viz., that to intend to do what one knows to be wrong is itself wrong.[11]

---

[10]Gauthier, esp. pp. 479-80; Fisher, p. 82. It should be noted that Gauthier's argument is in some ways the reverse of what I have represented in that he argues from the acceptability of the intention to the acceptability of retaliation itself.

[11]Kavka, 1978, p. 289. I critique Kavka's support for this principle in Chapter 5 (§4.2).

3. THE AIM OF THE THESIS

But there is another possible route to a moral defence of deterrence, which both accepts a (defeasible) version of the wrongful intentions principle and avoids the efficacy problems of the bluffing theories. This route proceeds from the claim that there are sets of circumstances in which an agent is justified in forming an intention which would be wrong to form outside that set of circumstances. Thus, while it may be true that the intention to retaliate is both required for effective deterrence and yet immoral because of the nature of what is intended, it is not the case that the deterring agent must be condemned for maintaining that intention. The purpose of this thesis is to carefully spell out this possible defence of deterrence, and to critique its plausibility.

To accomplish this task, I shall begin in Chapter 2 by setting out a few definitions and assumptions associated with the concept of nuclear deterrence, and then offer in Chapter 3 a roughly intuitive notion of ordinary intention against which we may compare and contrast deterrent intention. Before beginning to lay out the defence of deterrence, I shall in Chapter 4 look at the arguments against the necessity of forming a genuine intention to retaliate, arguments which take the form of suggesting various bluffing theories of deterrence. I shall show that at least one of these is plausible, and may reflect the actual attitude of those responsible for executing retaliation. Furthermore, I shall show in Chapter 5 that deterrent intention has been broadly misunderstood by critics on both sides of the moral question.

I shall then begin to examine this new defence by showing in Chapter 6 that the moral implications of the uniqueness of the intention have not yet been fully appreciated, and setting out in Chapter 7 a more accurate understanding of the dualistic nature of deterrent intention, i.e., that it comprises both the primary intention to deter

and the secondary intention to retaliate.

The general argument of Chapters 8 and 9, that there are situations in which an agent is justified in forming an immoral intention, takes the form of proposing and defending the Principle of Double Intention. As the name suggests, this principle is based on the Principle of Double Effect, a doctrine which serves as a yardstick for judging actions which result in both good and evil effects, where the agent intends only the good, but foresees the bad. In such cases, the agent may be judged to have acted morally, despite the evil which he produced. Similarly, the Principle of Double Intention is a tool for evaluating agents who intend good and intend evil within the same endeavour, i.e., as part of achieving the same overall objective. In these cases, the principle states that it may be morally acceptable for an agent to form both intentions, provided that, among other things, the primary or dominant intention is for the good.

The penultimate stage of the thesis then is an application in Chapter 10 of the general Principle of Double Intention to the specific question of nuclear deterrence. I shall argue that deterrence does in fact meet the criteria of Double Intention, and that as a result the deterring agent (i.e., the nation seeking to deter aggression by threat) may be justified in forming the immoral, embedded intention to retaliate.

In applying the principle to nuclear deterrence, we shall have assumed for the moment that a deterring agent can rationally form the secondary intention to retaliate, an assumption whose acceptance is as yet unwarranted. Thus, in Chapter 11 we shall return to confront the crucial issue which it must be said conceptually precedes any moral analysis of deterrent intention: Before we can determine if a deterring agent is morally justified in forming the retaliatory intention, we must first determine if it is even rationally possible for him to do so, given the admitted irrationality and immorality of retaliation itself.

Despite the normal conceptual sequence of rationality and morality, I have delayed discussion of the former because the issues of the morality of intention formation have a significant impact on the question of rationality. For if the argument of Chapters 9 and 10 is sound, the moral acceptability of the endeavour of deterrence (within the accepted limited context) provides the reasons which render forming the intention to retaliate rational.

The need both for development of the Principle of Double Intention and for a transposed analysis of rationality and morality demonstrates that, just as the advent of nuclear weapons necessitated a reexamination of the fundamental principles of international strategy and political theory, so it also necessitated a similar reexamination of the basic moral principles associated with intention formation. The *perestroika* within both disciplines has allowed for the acceptance of hitherto unthinkable phenomena regarding the conduct of war and national defence.

4. LARGER IMPLICATIONS

Given the vast array of possible methods for undertaking a moral examination of nuclear deterrence, it may be said that an in-depth study of intention as it relates to deterrence is so tightly focused that it runs the risk of missing the larger and more important moral issues of the policy. But the question of the morality of intention, while minute, goes straight to the heart of the larger question of the morality of deterrence as a whole. The moral uniqueness of the intention mirrors that of deterrence itself. Thus a careful dissection and evaluation of deterrent intention will provide the key to decipher the puzzle of the entire policy, and provide a demonstration of its moral justification.

Before completing these introductory remarks, we should note the continuing relevance of studies of nuclear deterrence even in the current age of political reformation within the Soviet Union and its allies. For it may be argued that policies of deterrence are now outdated; the sort of global threat to which deterrence once responded no longer exists. The unstoppable tide of quiet revolution has not only swept aside Communist domination in eastern Europe, it has in the process swept aside the need for nuclear vigilance. Unfortunately, that claim is premature. The need for deterrence policies will continue, for even if the Soviet Union no longer posed a major threat to the west, the danger that nuclear weapons may find their way into smaller but more radical hands may require the maintenance of a western deterrent.[12] Thus, analyses of deterrence will remain topical.

[12]The Lord Chalfont, chief British negotiator for the 1968 Nuclear Non-Proliferation Treaty, has argued recently that the imminent failure of the effectiveness of the treaty poses the single greatest danger to the West, and thus constitutes the most convincing reason to maintain a credible nuclear deterrent for the foreseeable future.

PART II: PRELIMINARY CONCEPTS

2: SOME DEFINITIONS AND ASSUMPTIONS

1. PRELIMINARY DEFINITIONS 14

1.1 Deterrence in General 14
1.2 The Policy of Nuclear Deterrence 17

2. SOME ASSUMPTIONS 19

2.1 Institutional Agency 19
2.2 The Acceptability of Just-War Theory 20
2.3 The Immorality of Nuclear War and the Escalation Hypothesis 21
2.4 The Deterrent Intention 23

3. TOPICS OMITTED 24

2: SOME DEFINITIONS AND ASSUMPTIONS

In keeping with the stated project of a critical analysis of deterrent intention, we begin with a short series of definitions of the concepts associated with the main topic. These are designed to clarify the meanings of crucial terms used in the thesis, and they begin to lay the foundation for the analysis which follows. Additionally, I shall make explicit several important assumptions about deterrence and nuclear war which are necessitated primarily by the limited scope of the work, and I shall also mention some of the areas which will not be examined.

1. PRELIMINARY DEFINITIONS

1.1 Deterrence in General

Of the two most important concepts in this thesis, 'intention' and 'deterrence', we shall reserve discussion of the former until Chapter 3, as it bears a significance which extends beyond the scope of the morality of nuclear deterrence. As to the latter, we may begin to clarify the concept by noting Edward Luttwak's remark that deterrence

is a species of 'suasion', a term he uses to depict the form of power which evokes in one's adversaries a positive (*per*suasion) or negative (*dis*suasion) attitude toward some contemplated action.[1] While what he calls 'compellence' is a type of persuasion which could be evoked in either one's friends or one's adversaries, deterrence, a version of dissuasion, can only be evoked in one's adversaries.

We may clarify this notion of suasion by introducing a definition of deterrence:

> *One deters when one successfully endeavours to prevent another from trying to achieve a particular goal by developing a barrier to achievement of that goal which is recognized as credible.*

The definition focuses on three essential elements of deterrence: First, deterrence is not a single action, but a complex policy designed to achieve an overall objective. As a result, it is more accurate to refer to deterrence as an endeavour, and not simply an act. Secondly, a deterring agent is attempting to achieve behaviour maintenance, not modification. He seeks to convince his adversary to maintain the relevant *status quo,* not to change it. Finally, the threat recipient plays an active role in deterrence. Deterrence is not accurately attributable to the agent who seeks to evoke that effect; it is more a quality of the respondent to that agent. Precisely speaking, one does not deter, one is (chooses to be) deterred, although the ordinary usage of the term tends to blur this distinction. In the realm of international relations in particular, as Luttwak notes, deterrence 'is not in the keeping of armed strength, but rather in the response of others to such strength."[2] This leads to two important points to note about deterrence. First, deterrence is a product of perception and belief: a potential adversary will *choose* to be deterred or not based on his perception of his threatening opponent. He will take into account such things as his opponent's ability to carry out

---

[1]Luttwak, p. 190.

[2]Ibid., p. 191.

the threat, the benefits and harms of acting despite that threat, and his beliefs about the resolve of the opponent.[3] I may seek to deter would-be intruders by placing a sign on my perimeter fence which states 'Beware of the Dog.' Whether or not the sign is effective depends not so much on my efforts, but rather on its perceived credibility. Absent the appropriate perceptions and beliefs, the mere existence of a 'deterrent' force cannot fulfill its purpose. The Japanese, for example, chose not to be dissuaded by the United States' decision (designed to induce dissuasion) to base their Pacific Fleet in Pearl Harbor in 1940 and 1941, and instead launched an attack against that force.[4]

The second point (which is also a lesson of this last example) is that the goals of deterrence cannot be assured simply by the deterrer maintaining the mechanisms of that policy. For this reason, deterrence is not inherently stable; its efficacy changes, subject to the changing perceptions and beliefs of the opposition. Deterrent effectiveness is a function of the credibility of the threat. A powerful deterrent force which is nevertheless perceived to be weak will fail to give an enemy pause; one that is excessively powerful, and perceived as such, will also fail to evoke the belief that the force will in fact be used.[5] Both extremes result in ineffective deterrence, and increase the possibility that the deterrent threat will be carried out.

Deterrence may be accomplished in two basic ways. The first is the standard method of deterrence by threat of retaliation. Sometimes referred to as punitive deterrence, it is the method of dissuasion most often thought of as deterrence, as

---

[3]A number of moral as well as strategic thinkers have commented on the prominence of belief in nuclear deterrence. Schelling (p. 36) notes that 'the threat's efficacy depends on the credulity of the other party,' while Kavka (1978, p. 287) more radically claims that 'deterrence depends only on the potential wrongdoer's *beliefs* of the sanction being applied.' For other views on belief see, e.g., Fisher, p. 79; Kenny, 1985, p. 79; Morris, p. 481; and Sterba. p. 101. The role of belief in deterrence will resurface during our discussion of bluffing in Chapter 4.

[4]Luttwak, p. 191, n.1.

[5]Walzer's example (p. 272) of a state which seeks to prevent murder by threatening to kill the family and friends of every murderer illustrates this last point.

exemplified by legal sanctions in a criminal justice system. The goals of the policy are sought to be achieved by announcing a credible threat to react to the occurrence of an adversary's unwanted act x by causing him to suffer costs which outweigh the benefits of doing x. If the deterrence policy is effective, the outcome of the adversary's cost-benefit analysis leads him to be deterred from doing x.[6]

The second method of dissuasion is deterrence by threat of denial. Here the emphasis is not on punishment after the fact, but prevention beforehand; burglar alarms and locks cause a would-be criminal to be deterred in this way. It is also the method employed by a conventionally armed defensive force. Although this seems to imply an odd use of the term 'threat', in that there is no reference to any potential infliction of harm (at least in the case of locks), a wider understanding of 'deterrence' includes the idea of prevention without an explicit reference to 'threat'.[7] And given that deterrence is a function of the recipient rather than the deterring agent, this form of deterrence clearly conforms to our notion. The goals of this policy are sought to be achieved by announcing a credible intention to prevent an adversary from accomplishing the results of x. As in the case of retaliatory deterrence, the adversary's cost-benefit analysis leads him to be deterred since his costs, no matter how small, will outweigh the now non-existent benefit.

1.2 The Policy of Nuclear Deterrence

The general concept of deterrence finds its most obvious application in the field of international relations since 1945, with the advent of nuclear weapons arsenals of

---

[6]For a discussion of the restrictions on punishment as an effective deterrent, see Honderich, pp. 58-60.

[7]Although some writers use the more restrictive definition, others understand deterrence in the broader sense. See e.g., Luttwak, pp. 199-200, Kemp, 1987a, pp. 278-79, and especially Hardin, pp. 187-88, who, while noting that 'etymologically, to deter is to provoke terror,' points out that traditional, conventional deterrence has always been based on denial rather than punishment.

first the United States and then the Soviet Union. The policy for both of these countries, as well as for more recent members of the 'nuclear club', has been to emphasise the arsenal as a foundation of retaliatory deterrence. Each country has explicitly (through public policy statements) and implicitly (through preparation and training to employ the weapons) declared its willingness to react to its adversaries' aggressive behaviour by launching a nuclear attack aimed at inflicting unacceptable damage to his vital interests.[8]

While retaliatory deterrence is the usual form in international politics, dissuasion by denial is also possible. Here one may think of the United States' Strategic Defence Initiative to develop a means of counteracting the effectiveness of ballistic missiles, as well as the less prominent concept of preemptive deterrence designed to remove a potential threat of aggression before it is actualised, a policy which the Israelis are often accused of implementing, but one which has also been suggested for the United States to pursue with regard to regional dangers to its vital interests.[9] But since the moral issues are by far more serious for retaliatory deterrence than for deterrence by denial, and since the actual international political situation currently and for the imaginable future is bound up with retaliatory deterrence, we shall be concerned here only with that type.

---

[8]The United States' objective in deterrence, which reflects that of its western nuclear allies Great Britain and France, was clearly formulated in the 1966 defence budget statement and has not been substantially altered: nuclear war forces are in place 'to deter a deliberate nuclear attack upon the United States and its allies by maintaining a clear and convincing capability to inflict unacceptable damage on an attacker, even were that attacker to strike first.' (Senate Armed Services Committee and Senate Appropriations Committee, *Military Procurement Authorization, Fiscal Year 1966*, p. 43, quoted in Ball, p. 69). Soviet deterrence objectives are relatively the same, although their targeting strategies differ fundamentally from many western powers. See Lee, esp. pp. 97-99.

[9]See Nacht, p. 20: 'The United States should seek to define the necessary and sufficient conditions for engaging in acts of preemptive deterrence. The purpose of such acts would be both to protect assets vital to American interests and to dissuade the Soviet Union from attempting to seize these same assets.'

2. SOME ASSUMPTIONS

Because this thesis is by nature limited in scope, there are a number of subsidiary topics the examination of which, although important and illuminating, would lead us well away from our main concern. However, it may be wise at this point to mention a few of the more significant of these diverse issues, and to make explicit the assumptions we shall hold regarding conclusions about them.

2.1 Institutional Agency

When one begins to probe the question of the morality of deterrence, one must soon confront the prior issue of whether a corporate body of individuals--a nation-state in this case--should be treated as though it were an individual moral agent. This issue informs the question of who bears the moral responsibility for the actions of the state. The first answer to this question, that all legitimate members of the state, i.e., its citizens, are jointly responsible, is grossly counter-intuitive. Even in an ideal democracy, newborn babes-in-arms surely cannot be held accountable for the sins of their government. Similarly, the next logical answer that the voting members (of a representative democracy) share the culpability of their elected leaders, is problematic. Is a voter whose candidate loses nevertheless answerable for his leaders' actions?

The issue of corporate responsibility becomes more perplexing when the problem of intention formation is introduced, as it will be in the present work. Regardless of the level of consensus, a state cannot be said to have a mind of its own. But if intention is a mental state, whose mind is relevant in the ascription of intention to a nation, or to any corporate body? Or is it that institutions are incapable of intending?

These questions are at once puzzling and alluring, but must be bypassed in favour of an unfortunately bland but acceptable assumption that we may allow a limited concept of institutional agency, and may roughly ascribe intentions to a state, or (if this causes undue discomfort to strict realists) at least to the leaders of that state directly responsible for implementing and executing the retaliatory nuclear deterrent. This assumption will at least permit us to hold nation-states morally accountable for their intentions regarding nuclear deterrence.[10]

2.2 The Acceptability of Just-War Theory

Underlying any discussion of the morality of nuclear deterrence must be the moral assessment of war in general and nuclear war in particular, since deterrence (the threat of war) is conceptually related to war in an intrinsic way. The most readily available yardstick for determining the morality of war is the tradition of moral precepts which have evolved into Just-War Theory.[11] While some of the seven criteria are subject to controversial interpretation,[12] the two which will concern us, discrimination and proportionality, are relatively settled. A nation which engages in war (or in a particular act of war--these two criteria may be used to establish either *jus ad bellum* or *jus in bello*) will satisfy the criterion of discrimination just when it has made every (reasonable) effort to ensure that noncombatants (and ex-combatants) are not put at risk, that is, when it has complied with the provisions of the Geneva Convention 'to alleviate

---

[10]For a more extensive discussion of this issue, see e.g. Shaw, p. 256; Hardin, pp. 169-72; and especially O'Neill, pp. 57-67, where she argues convincingly for the claim that institutions are moral agents 'as defined in terms of the cognitive capacities and powers of action available in a given context.' (p. 66).

[11]To cite only two of the many fine works on Just-War Theory, James Childress (pp. 427-445) provides a succinct and well-researched examination of the theoretical underpinnings and purposes of the criteria of the theory; Kenneth Kemp (1987b, esp. pp. 62-90 and 110-29) expands Childress's findings, and applies them to the particular moral problems of war.

[12]For example, the question of who constitutes a 'legitimate authority' is particularly troublesome in the case of civil war or insurgency. See Kemp, 1987b, pp. 112-113.

the sufferings [of civilians] caused by war."[13] Clearly, the direct intentional attack of civilians is prohibited.

The second criterion is designed to ensure that the good which a nation hopes to achieve by going to war proportionally outweighs the harm which must necessarily accompany that endeavour. As Paul Ramsey writes, 'It can never be right to resort to war . . . unless one has reason to believe that in the end more good will be done than undone or a greater measure of evil prevented."[14] The correct mechanism for comparing harms and benefits is open to dispute, but we may say without much fear of contradiction that the prospect of a vast number of civilian deaths will weigh decisively against any operation which would produce such a result.

2.3 The Immorality of Nuclear War and the Escalation Hypothesis

Acceptance of these two criteria as at least a partial basis for the moral assessment of war leads us to the conclusion that any launch and detonation of nuclear weapons is morally forbidden. And for the purposes of the arguments contained in this thesis, we may accept the assumption that this prohibition is absolute; nuclear attack may not be considered an option for resolving potential or actual conflict.

Many of the standard arguments purporting to demonstrate the immorality of nuclear deterrence rely heavily on the claim that any detonation of a nuclear warhead is immoral. This claim is in turn based on the Just-War argument which concludes that global nuclear war is immoral (since it disproportionately and indiscriminately annihilates innocent civilians), plus a claim which may be called the escalation hypothesis, that in any conflict, once the nuclear threshold is crossed, that is, once

---

[13] *Geneva Convention Relative to the Protection of Civilian Persons in Time of War*, Article 13.

[14] Ramsey, p. 195.

a nuclear weapon is used in combat, the conflict will escalate into an all-out exchange between the nuclear powers. Each will launch his nuclear arsenal at his opponent, reeking devastation not only on that opponent, combatant and noncombatant alike, but also on a significant portion of neutral territory, due to the nature of fallout from nuclear warfare and the catastrophic climatic effects of even a comparatively small number of nuclear detonations. Indeed, this quite possibly might signal the end of the biological dominance of man on Earth.[15] Because even one such explosion will lead to this catastrophe, any wartime use of nuclear weapons is deemed immoral.

But before accepting the escalation hypothesis, we should note that many strategic thinkers doubt the veracity of the claim. 'Escalation is not a mechanistic process to the outcome of which no human agent can contribute after the initial decision.'[16] Escalation is not inevitable. Once the nuclear firebreak has been crossed, it might well be the case that the parties to the conflict (assuming some degree of rationality) will see that escalating the destruction is purposeless. Each would see that it would be in his best interest to terminate the escalation, if not the entire conflict, as rapidly as possible. Escalation which leads to global devastation, while a possibility, should not be considered inevitable.

The questionable acceptability of the escalation hypothesis will become pivotally important in the argument (in Chapter 11) for the rationality of forming the intention to retaliate. If one accepts the inevitability of escalation, then it will be the case that forming the intention is not rationally possible. Alternatively, if one denies the

---

[15]For detailed analysis for the effects of nuclear war on the global ecosystem, see Schell, esp. 71-96; and Sagan, pp. 250-55.

[16]Fisher, p. 99. See also pp. 96-105, where he carefully lays out the arguments for and against several versions of the escalation hypothesis, eventually favouring a rejection of it. Luttwak (pp. 120-23) also doubts the inevitability of escalation, as does Schelling: 'The victim [of a nuclear explosion] may not see wisdom in unleashing all-out war in response to the pain and insult.' (p. 195). Against these views, Finnis & co. (esp. pp. 148-49) argue that not only is escalation inevitable, but it also plays an essential part in effective deterrence, which must rely on the threat of final retaliation involving the launch and detonation of a nation's entire nuclear arsenal. For additional support for the hypothesis, see also Kenny, 1985, pp. 27-31.

hypothesis, the rationality of the intention is resolved, and with it the entire proposed moral defence of deterrence.

Since escalation is only tangentially relevant to the majority of the arguments presented in this thesis, I shall accept the hypothesis that escalation will take place and the resulting conclusion that any use of nuclear weapons is immoral. But at the critical juncture in Chapter 11, I shall assent to the conclusion of Fisher, Luttwak and Schelling that the hypothesis may be abandoned with good reason. Thus my defence of deterrence will be limited to circumstances in which the hypothesis of escalation can be justifiably rejected.

2.4 The Deterrent Intention

Deterrence involves dissuading an adversary from certain activities by threatening him with unacceptable consequences should he decide to ignore the warnings. In order for deterrence to be effective, the announced threat must be credible, which generally means that it must be founded on a publicised intention to carry out that threat. It is this intention that demonstrates the commitment necessary for credibility.[17]

John Finnis & co. go to some length to argue that the intention to retaliate is a necessary part of effective deterrence.[18] For the most part, their argument seems sound, although we shall have occasion in Chapter 4 to challenge their conclusions about the feasibility of bluffing as an alternative. Suggestions that something less than a full-blown retaliatory intention would suffice (e.g., a willingness or readiness to respond), while perhaps strictly accurate, do not seem to make much difference to the moral assessment of deterrence. As Anthony Kenny notes, 'If it is true that it is

---

[17] Schelling, p. 36.

[18] Finnis et al., pp. 104-31.

wrong to intend what it is wrong to do, it is equally true that it is wrong to be willing or ready to do what it is wrong to do.'[19] Barring the possibility of a deterrent bluff, any realistic examination of nuclear deterrence must accept that the intention to retaliate is a genuine part of the enterprise. Thus we shall assume the necessity of that intention, at least until we have the chance to examine the alternatives in Chapter 4.

3. TOPICS OMITTED

As must be the case with any subject whose implications and applications are as far reaching as those touching upon nuclear deterrence, we must limit the scope of our examination. To do so, I shall, in addition to focusing this thesis by accepting the assumptions mentioned above, pass over a number of related and ancillary topics, many of which involve the investigation of empirical questions or questions of purely strategic political or military interest, such as whether nuclear deterrence is in fact effective in preventing aggression, and whether deterrence is the only method of prevention.[20] Additionally, I shall not examine the moral dimensions of extended (i.e., designed to include the protection of allies) versus minimum deterrence, or of strategic versus tactical nuclear weapons, since neither of these issues significantly affects the deterrent intention. Nor shall I discuss the moral impact of a 'first-strike' capability and the related question of strategic defence. Finally, I shall only briefly mention (in Chapter 12) the effect on nuclear deterrence of what may well become the most significant and dramatic series of events of our time, the (largely non-violent) political

---

[19] Kenny, 1985, p. 50.

[20] Despite their eventual condemnation, Finnis & co. (esp. p. 77) argue that deterrence is the only way for the western nations to secure peace and security; others, such as Hedley Bull (p. 14), question the certainty of that conclusion: 'Was it, in fact, the prospect of assured destruction or other factors--memories of World War II, fear of domestic turmoil, fear of economic dislocation--that led the Soviet Union to conclude that any attack was not worthwhile?'

upheavals in eastern Europe and the resulting dissolution of the 'Iron Curtain'.

While many topics must be left untouched, there is much of importance in what remains. For the larger question of the morality of nuclear deterrence is centred on the nature of the intentions formed by the deterring agent.

3: A PLAUSIBLE NOTION OF INTENTION IN GENERAL

---

1. ORDINARY INTENTION 28

2. THE LINEAR RELATIONSHIP MODEL 30

   2.1 The Role of Desires and Outcomes 31

3. OBJECTIONS TO THE MODEL 35

   3.1 *Akrasia* 35
   3.2 The Problem of Conditional Intentions 37
   3.3 Intention-Action Causality 42

4. OBJECTION TO INTENTION-ACTION INDEPENDENCE 44

5. INTENTION AND NUCLEAR DETERRENCE 47

3: A PLAUSIBLE NOTION OF INTENTION IN GENERAL

What kind of super-strong connexion exists between the act of intending and the thing intended?

--Ludwig Wittgenstein

A careful and precise study of deterrent intention must include a basis for comparative analysis in the form of an acceptable notion of intention *simpliciter.* This is especially true if one wishes to claim, as I do, that deterrent intention differs significantly from ordinary intention.

Thus the general purpose of this chapter is two-fold: First, I shall very briefly set out a rough, but intuitively acceptable notion of intention in general. Because anything other than a cursory account of the concept would require us to depart the realm of moral philosophy and venture deeply into the philosophy of mind, this sketchy notion will admittedly fall well short of constituting a fully defensible account of intention, leaving untouched many of the conceptual issues lingering around the concept.[1] Despite this limitation however, I hope to present a plausible picture of

---

[1] A more thorough discussion of these questions can be found in Anscombe, 1966, esp. §§9-23 (although in some sense she too fails to provide a *bona fide* definition); and more recently in Davidson, 1980; and Bratman, 1987.

intention which will be adequate to show that *deterrent* intention differs from the ordinary sort in ways which are important enough to affect our moral judgement of the deterring agent.

Secondly, I shall outline an argument for the claim that intention can be conceptually divorced from action, and is thus a proper subject for independent moral evaluation. The argument will take the form of answering Wittgenstein's question by providing a description of the relationship between intention and action designed to demonstrate that the two, although intimately connected, are not therefore indivisible.

1. ORDINARY INTENTION

As it is commonly used, the meaning of 'intend' in an expression such as 'I intend to see *Hamlet* at Stratford tonight,' does not seem to be problematic. Intention statements express a commitment by the intending agent to act. This commitment arises from the agent's recognition of a preference either to fulfill a desire or to accept an obligation to act. The formation of an intention leads to action the result of which (all things being equal) is satisfaction of one's preferences. To put it another way, intentions serve to crystalise preferences into plans.

This vague description of how intention leads to action stands in need of clarification. But before we can begin that task, we need some idea of what an intention is. Formalising the intuitive idea of intention, we get our first attempt at a definition: *An agent intends an action if he (1) knows he is doing it and (2) wants to do it for its own sake or for some other end.*[2] I intend to see *Hamlet* because I know that I am doing so, and I want to do so for the enjoyment watching it brings to me.

---

[2]This definition is derived from Kenny, 1975, p. 56.

Thus this definition is a good start. But while it captures the majority of intentions, a literal reading of the conditions omits at least two important types of genuine intending. First, since it considers actions which are performed intentionally, and not intentions themselves,[3] intentions formed about future actions seem to be ruled out by (1). Because an agent must know that he is doing x in order to intend it, he cannot, strictly speaking, intend to do x at some point in the future. I cannot, for example, intend to see *Hamlet* until I am actually doing so. Taken literally, this condition thus eliminates many, perhaps even most, seemingly genuine intentions, which except in the simplest cases involve future action. So we may broaden the definition by allowing for either a more liberal interpretation of (1), or else modify it slightly so that having an intention requires that the agent know that he is *able to* perform the act intended. Emphasising ability rather than contemporaneous performance, this modification includes all cases of intention already embraced by (1), but also allows for the inclusion of future-directed intentions.

The second group of intentions omitted by the definition results from the overly restrictive set of reasons for intending listed in condition (2). Although it lists both instrumental and intrinsic desires, it overlooks those types of intentions which an agent may form as a result of his recognition of some obligation to act, without any (at least conscious) view toward fulfilling a desire. One may feel that it is one's duty, for instance, to keep a promise, despite any inclinations to the contrary (the core of Kantian morality). So in order to gather all reasons for intending within this condition, we may substitute 'prefers' for 'wants', since an agent's preferences more accurately reflect the outcome of practical reason which results in intention formation, and thus

[3] To use Anscombe's distinctions among the uses of 'intention' (1966, §1), the definition addresses 'acting intentionally,' not the 'expression of an intention.' We shall return to these distinctions in Section 4 below.

have a wider scope than do his mere desires.[4]

So the modified definition of intention reads: *An agent intends an action if he (1) knows he is able to do it and (2) rationally prefers it over any alternative actions.* Although by itself, the definition cannot provide a comprehensive analysis of ordinary intention, it is complete enough to provide the required baseline for comparison with deterrent intention which will serve (in Chapter 6) as a demonstration of the uniqueness of that type of intention.

2. THE LINEAR RELATIONSHIP MODEL

Armed with this definition, we may begin to analyse the connection between intention and action. The claim which I wish to advance here is that, at least for the purposes of moral analysis, intention is conceptually independent of action. I shall support this main claim by suggesting a picture of the relationship between the two as a linear relationship, viz., forming the intention to do x leads an agent directly to the performance of x, barring any change in the relevant *status quo* between the time of the formation and the time of the performance. The agent forms the intention as a step towards linking his preferences to the end result of the action; preference, intention, act and outcome together form the linear progression of the action sequence.

Obviously, the depiction of this model is not meant to constitute a formal proof of the claim that intention and action are conceptually independent. It is meant rather to offer more intuitive support for that claim, and against the opposing view that intention is part of action.

---

[4]Unfortunately, this is not the forum to debate to the soundness of Kant's moral philosophy, especially as it contrasts with the Hobbesian idea that morally right actions are in the final analysis egoistically grounded. Even if the Hobbesian view is correct, it is for our purposes more perspicuous to modify condition (2) to include recognition of duty as a motivation for intention formation.

2.1 The Role of Preferences and Outcomes

To see the acceptability of the model, we may expand our examination beyond intention and action to include preferences and outcomes as well. From whence do intentions arise? The obvious (and it seems correct) answer is that they are generated from the agent's preferences (i.e., either his desires or his recognition of duty). It is his ordering of his preferences which leads an agent to formulate plans of action. Preference is the beginning of the action sequence, and, in the final analysis, the ultimate answer to the 'why?' of action. The intention-action linear relationship envisaged by the agent has its genesis here. As we have noted, the intention to do x springs from the agent's realisation that he can do and prefers to do x. This part of the intention picture is reflected in Anscombe's understanding of Aristotle's practical syllogism:

> Aristotle would seem to have held that every action done by a rational agent was capable of having its grounds set forth up to a premise containing a desirability characterisation; and as we have seen, there is a reasonable ground for this view,[5]

although, as we have seen, desirability may not provide a sufficiently wide scope for intention formation.

There is a further aspect of the notion of intention which is especially relevant to the subsequent discussion of deterrence, that is the relationship of outcome to intention. Following Kenny (and thus Von Wright),[6] we can distinguish between several types of outcome. A *result* is the 'upshot of the change by which an act is defined.' My act of traveling to see *Hamlet,* for example, results in my arrival at the Royal Shakespeare Theatre in Stratford. A *consequence* is any other subsequent

---

[5] Anscombe, 1966, §38.

[6] Kenny, 1966, p. 645.

change, such as my return home. A *goal* is the ultimate end for which an action is undertaken. In this example, that would be the enjoyment of the play, although the same action (and result) might lead to someone else achieving a different goal, say a critical review of the play or the fulfillment of some promise to attend. In the simplest instances, the goal will be identical with the result. More complicated cases (which I shall call endeavours) will include *means*, intermediate steps (both actions and results) which lead to achieving the goal. Purchasing tickets is a means to my goal. Finally, *side effects* are those outcomes which are neither consequences nor means in the endeavour to achieve the goal. Here we might imagine the death of the bugs which are unfortunate enough to end up on my windscreen as I drive to Stratford.[7]

Side effects are distinguished from means in that they are neither wanted nor sought after by the agent; they are distinguished from consequences only by the fact that they do not follow the goal (or result), either temporally or causally. For that reason, the moral assessment of consequences and side effects is virtually the same, while the moral assessment of means differs greatly, although it is usually tied to the assessment of the goal, since deciding upon (i.e., forming an intention to achieve) a goal includes defining the means necessary to achieve it. This separation of means and goals from side effects and consequences is the moral justification behind the Principle of Double Effect, discussed in Chapter 8.[8]

Although actions, and not results, are the proper objects of intentions, the linear relationship that exists between preferences, intentions and actions can be seen to extend through actions to results. To say 'A intends g,' where g is the goal of doing

---

[7]While this may be considered a nonstandard use of the term of 'side effect,' especially in medical contexts, I have retained it to highlight the difference between consequences (i.e., unintended outcomes following the action) from side effects (i.e., unintended by-products which do not follow completion of the act).

[8]The stipulated meanings of the terms may be disputed (see, e.g., Hart, 1985, pp. 27-28, where 'result' is 'the culminating phase' of a consciously designed process, i.e., a goal in Kenny's lexicon), but the point of their technical usage here is meant to underscore a narrower, more well-defined extensions of the various types of outcome.

x, is in fact an elliptical version of the fuller and more accurate intention statement, 'A intends to do *x* in order to (achieve) g.' That this relationship may not always be obvious is primarily due to the shorthand convention used in many intention statements. When asked about his intentions, an agent will probably answer by mentioning the goal he hopes to achieve, and leave implicit any discussion of the action itself. This would be especially true of statements about future-directed intentions where the act is not yet, or cannot be, fully specified, where 'to do *x* in order to . . .' is replaced by 'to act in such a way as to . . .' This would be the case for an agent who has identified a preference to be fulfilled, but has not (yet) formulated his plan for doing so.

Obviously, there are a number of factors, both internal and external to the agent, which have a bearing on whether an action will in fact follow from the formation of its intention. A change in external factors which interrupts the linear flow from intention to action will (usually) prevent an agent from acting, *despite* his intention. It might well be that I have formed my intention to see *Hamlet*, unaware of a change in the theatre schedule. In this case, my intention will be unfulfilled, but of course will still be an intention. Interrupting factors internal to the agent most often take the form of an abandonment of the original intention, either purposefully (a change of mind) or inadvertently (forgetting the intention). On the way to the theatre, I may decide (hopefully for some justifiably good reason) to go somewhere else instead. In this case, my intention is not unfulfilled, but rather abandoned. Here it is natural to speak of intentions in the past tense: 'I had intended to go, but . . .'

So it is that, knowing the intentions of an agent, one looks for a reason why the associated action was not carried out. It would be odd indeed if, knowing that 'A intended to do *x*,' and also that '*x* was not done,' we could not find some factor which changed the relevant *status quo*, and caused an interruption in the expected linear

flow.[9] In fact, one would probably be forced to admit in that case that the (true) intentions of the agent were not really known, that 'A intended to do x' is false.

The linear relationship model begins to give us a picture (albeit sketchy) of the sequence of preference, intention, action and result: An agent forms (or recognises) a preference, and determines to fulfill it. This moves him to formulate a plan of action and the intention(s) to carry out that plan. The execution of his intention(s) produces a result. By linking his preference to the result, the agent forms his intention to act so as to realise that (now intended) result. In ordinary circumstances, the entire linear process is considered successful when the ultimate goal 'matches' the initial preference, that is, when the agent achieves what he had set out to accomplish. As I sit in the balcony at the Royal Shakespeare Theatre and watch the curtain fall on the sight which 'shows much amiss' in the great hall of Elsinore, I may reflect, with some satisfaction, that I have achieved that which I set out to do.

The linear linkage of preference, intention, act and result is most apparent in dynamic situations where the agent must change his intention (and his action) in order to 'track' his goal. Suppose for example that I want to attend a certain conference in Blackpool next month. I develop the intention to go to Blackpool, and begin to act, developing travel plans, etc. However, before departing I discover that the conference venue has been changed to London. Since I am tracking my preferred goal (attendance at the conference) rather than any incidental consequences or side effects (such as a visit to Blackpool), I change my plans and formulate a new intention to travel to London. Thus I adjust my intention(s) and actions in order to achieve that successful match of preference and goal. The consequences and side effects of my plans, e.g., the chance to visit Blackpool, are disregarded because they do not stand

[9] Kenny (1973, p. 131) makes this same point: 'If a man intends to X on occasion C, and does not X on C when he can do so, some explanation is called for.'

in that special linear relationship to my preferences as do the results.

3. OBJECTIONS TO THE MODEL

There are three objections which may be raised against the linear relationship model. The first may be called the *akrasia* objection, the second concerns conditional intentions, and the final objection offers an alternative to the model which depicts a causal relationship between intention and act. I shall examine each in turn.

3.1 *Akrasia*

Suppose that Mr Smith, an alcoholic, forms the intention to stop drinking, but does not stop, despite his continuing protestations that he still intends to do so. By the linear relationship model, intention leads to action unless the agent is prevented from execution, or else abandons that intention. In the case of *akrasia,* it seems that we are forced to conclude either that the agent never really forms the intention, or that he has changed his mind and maintains the intention no longer. Does Mr Smith actually intend to stop drinking? He certainly says that he does. But yet he continues to drink. Can the linear relationship model explain this?

Before answering this question, we need to be clear on what *akrasia* means. After concluding to act as the result of practical reasoning, there seems to be a spectrum of weaknesses to which one can fall victim: If an agent expresses an intention, but takes no steps whatsoever towards fulfilling that intention, then his will is not so much weak as it is nonexistent. The real difficulty arises when an agent takes some steps toward fulfillment but stops somewhere short of achieving his goal, and this cessation cannot be explained by either prevention or abandonment of the intention.

Perhaps the problem lies in our limiting the interruption factors to two. Can *akrasia* be a third? If so, then uttering an intention expression, but not acting on it, can result from three causes: (1) prevention; (2) abandonment; or (3) *akrasia*. The cause of Smith's drinking would be (1) if, for instance, Smith's old drinking mates, angered by his apparent self-righteousness, spike his tea, or waylay him and force a pint down his throat. The cause would be (2) if Smith had actually intended to stop drinking, but now decides to abandon his intention and have a drink. The third cause would arise if Smith (actually) intends to stop drinking, but cannot help himself when he is near the bottle.

But attributing the interruption to *akrasia* presents some difficulty. What is the status of his intention while he is drinking? Has he abandoned it temporarily, or somehow suspended it, or merely forgotten it? This last possibility seems unlikely, since (we would guess) that simply being reminded of his intention during a drinking binge would probably not cause Smith to say, 'Oh, that's right. I forgot,' and stop drinking immediately. But then how is temporarily abandoning or suspending that intention any different from cause (2) abandonment? Maybe there is no need for a third cause.

However, the need may be clearer if we consider the case of Mr Jones, who suffers from claustrophobia, and is about to enter an elevator to ride up to the 17th floor. As the elevator approaches, Jones forms the intention to get in, even though he knows of his debilitating fear of such enclosed spaces. The doors open, Jones hesitates, but steels himself and moves forward. As he approaches the entrance however, he stops, and does not go in, despite knowing that his fear is irrational. Here it seems that his intention is not strong enough to motivate action, that *akrasia* has prevented him from fulfilling his intention. While it may be debatable whether *akrasia* (in this form at least) should be considered a separate category of impediments to

action (one may want to argue, for instance, that phobias should really be classified as external preventers of action[10]), it is clear that it represents at least a subgroup of such impediments, and as such may be accepted within the linear relationship model.

3.2 The Problem of Conditional Intentions

The second objection to the model has to do with conditional intentions, those of the form, 'A intends to do *x* on the condition that C occurs,' where C is (at least partly) outside of the agent's sphere of control. To expand on our earlier example, it might be that I intend to see *Hamlet* tonight if you accompany me. In cases such as this, it seems that the linear chain is broken, that the flow from intention to action is significantly altered from the earlier, simpler case. Instead, it seems that the action bears only a contingent relation to the intention, viz., contingent upon the fulfillment of the condition(s).

For reasons which I shall explain shortly, the differences between conditional and ordinary intentions do not affect the plausibility of the picture of intention which I have so far sketched. But I do think it important to examine, and refute, the commonly held view that there is no morally significant difference between the two types.

There are basically two reasons why someone would claim that there is no difference. The first would be to argue from the premise that all intentions are in some sense conditional. That is, in order for an intended action to be carried out, all relevant conditions, most of which are implicit and internal to the agent, must obtain.[11] These

---

[10] It may also be argued that irrational acts based on phobias are not genuine cases of *akrasia*, since they are not intentional, a condition which Davidson (p. 21) and his critics (e.g., Peacocke, p. 52, and to some extent Grice and Baker, p. 48) all agree is essential.

[11] John Lango (p. 319) refers to these conditions as 'requisites'.

include, but are not limited to, the appropriate occasion, the agent's correct mental state, a lack of external impediments, and, finally, other conditions (if any). My earlier (unconditional) intention to see *Hamlet,* though not explicitly stated as such, was conditionally based on the theatre schedule, availability of tickets, transportation, my continuing desire to go, and other factors. Although it is the last category of pre-requisites which usually determines if an intention is classified as conditional, all intentions are conditional in the sense relevant to the objection being raised against the linear relationship view of actions and intentions.

This line of reasoning, however, fails to support the claim of equivalence. The 'conditions' of ordinary intentions do differ from the genuine conditions of conditional intentions in that the latter are not merely *presupposed* in the background of practical reasoning, but play a substantive role in the fulfillment of the intention, or as Davidson puts it, 'are reasons for acting that are contemporary with the intention.'[12] Ordinary conditions do not play such a role. My intention to communicate the ideas in this thesis presupposes, for example, that you, the reader, are literate. But I do not therefore (in any ordinary sense of the term) conditionally intend to so communicate. The conditions of ordinary intentions are at best only important in (sophistical) philosophical arguments.

A further significant difference between the two types of conditions lies in the fact that the ordinary type, if they are more than merely background presuppositions, are within the sphere of the agent's control. This is after all what it is meant by the first part of our earlier definition of intention, viz., that having an intention requires that the agent know that he is *able to* perform the act. The force of an ordinary intention is derived from the commitment which the agent demonstrates in recognising that he has the power to realise his preferences. Genuine conditions, on the other

---

[12]Davidson, p. 94.

hand, are usually predicated because they involve states of affairs which are outside the agent's control. This is why my intention to see *Hamlet* if you go with me is conditional in a way in which my original intention is not, despite the conditions mentioned above. The force of the intention is altered by the agent's dependence on external circumstances.

There is a second line of reasoning, presented best by Finnis & co., which can be made against distinguishing conditional and ordinary intentions.[13] The primary purpose of the study of intention, at least for moral theorists, is to gain an understanding of the agent's state of mind, which in turn is an aid to ascribing moral praise or blame. This is the point, for example, behind the wrongful intentions principle ('To intend to do what one knows to be wrong is itself wrong.'), which seeks to link the intended (but perhaps unfulfilled) actions of an agent to his moral character.[14] Given this point of view, the existence of a set of conditions as a prerequisite for action has little or no bearing on the state of mind of the agent, or more to the point, our moral assessment of that state of mind. Sirhan Sirhan's intention to assassinate Robert Kennedy in cold blood was indicative of his relevant state of mind, and consequently his moral character, whether or not that intention was contingent upon Kennedy, after his victory speech, exiting his hotel via the kitchen facility. It is only the execution of the intention, and not the intention itself, which is contingent upon the fulfillment of any conditions. As Finnis & co. put it, 'conditional intentions are not conditional in so far as they determine the self, but only in so far as outward behaviour is still to be determined by them.'[15] For the purposes of moral judgement, this character-shaping aspect of intention formation (which can also be found in

---

[13] Finnis, et al., pp. 81-83.

[14] I discuss the principle at some length in Chapter 5 (§4.2).

[15] Finnis, et al., p. 82.

ordinary intentions[16]) is critically important. We are justified (all things being equal) in condemning an agent who forms the intention to act immorally, regardless of whether the execution of that intention depends upon circumstances which are presumably outside his control. Thus, whether or not an intention is classified as conditional has no bearing on moral assessment.

But the Finnis argument is wrong for this reason: Moral justification may very well turn on the nature (and likelihood) of the predicated condition. That is, whether I am justified in intending *x* if C depends on whether doing *x* in C is justified, even if doing *x* *simpliciter* is not, as well as the probability that C will in fact occur. In normal circumstances, failing to stop at a red light is wrong, but it may be justified if the driver has sufficiently good reason for doing so (e.g., an emergency) and if he ascertains that there is no conflicting traffic ahead. The conditions have a significant impact on our moral assessment of the intention because they fill in the details of the intention. Surely there is a moral difference between a man who intends to kill another, and a duly appointed executioner who conditionally intends to kill a prisoner *if* the prisoner is found guilty by due process, and has exhausted all appeals, etc.[17] Yet if we discount the conditions, it is difficult to see what would distinguish the two would-be killers. It would not help the Finnis case to argue that the act descriptions are different, and therefore their intentions are not the same, since the only relevant descriptive differences lie in the conditions of execution (so to speak). To rely on those differences is to admit that the conditions are in fact morally significant.

It is true that this reading of the moral difference between the two types depends on the assumption that carrying out the act is at most *prima facie* wrong. It would seem

---

[16]See, e.g., Wittgenstein (§659), who writes that intentions tell 'something about *myself*, which goes beyond what happened at that time.'

[17]I am grateful to Anthony Kenny for suggesting this example.

to be another matter if the act in question were absolutely prohibited, for that would mean the act is unconditionally ruled out, that it is not an option, regardless of any conditions which may affect its execution. And in fact Finnis & co. are concerned only with the conditional intention in deterrence involving an act (i.e., intentional killing of the innocent) which (they argue elsewhere[18]) is absolutely prohibited.

However, even with respect to absolutely prohibited acts, Finnis & co. have not provided a convincing argument for the homogeneity of ordinary and conditional intentions. First, they have failed to prove an absolute version of the wrongful intentions principle, viz., that it is *absolutely* wrong to intend that which it is absolutely wrong to do. As they show in their lengthy argument for the plausibility of the common morality (and for the wrongful intentions principle), acts are wrong in so far as they 'destroy, damage, or impede some instance of a basic human good."[19] Since intentions do not do so to the extent that actions do, it is not readily clear that the absolute prohibition can transfer. And such a proof is by no means obvious, given the problems of the defeasible version of the principle which I discuss in Chapter 5. Even that version, if acceptable, would only allow that intentions to commit absolutely immoral acts are at most *prima facie* wrong. And this of course is just what the above distinction between conditional and unconditional intentions requires. Secondly, they have not sufficiently accounted for the morality of intending acts which are unlikely to occur. An agent who intends a praiseworthy but unlikely act is less worthy of approbation than one who intends the same act but is in a position to carry it out. The same comment can be made about conditional intentions. A corporation which announces its intention to donate a large sum of money to aid famine relief, provided that the amount is matched by public donations is certainly worthy of commendation, but not

---

[18]Finnis, et al., esp. pp. 297-300.

[19]Ibid., p. 293.

to the degree of a company who unconditionally intends to donate the same amount.

There is, then, a morally significant difference between conditional and unconditional intentions. But the existence of the former does not impugn the plausibility of the linear relationship model which I have described. As a picture of how intention leads to action, the model is compatible with conditions which are attached to execution of the intention. The agent's formation of his plan of action, which includes intention formation, is still an attempt to match his preference to the goal, even when the intentions formed are conditional on the fulfillment of one or more external requisites. The linear relationship between act and intention is still envisaged by the agent; he still seeks to carry out his plan to match his preference with his expected outcome.

Additionally, it must be remembered that the model is in part designed to show that intention is independent of action. That some intentions may not lead to execution without the fulfillment of conditions can only serve to underscore that independence. As a result, the existence of conditional intentions does not adversely affect the suitability of the linear relationship model, at least for our present purposes.

3.3 Intention-Action Causality

The final point which we consider in connection with the model is an alternative interpretation which views the relationship between intention and action as a causal one: The way in which intention leads to action is that it *causes* action. This interpretation has at least some intuitive appeal, since intentions do seem to produce or bring forth action. But despite this appeal, the view suffers from either inaccuracy or deficiency. In one sense, it is inaccurate to say that intention causes action. For if it were true, then intention without action, e.g., unfulfilled intention, would be

impossible, since intention *qua* cause must (all things equal) produce action. Once an agent formed an intention, the intended action would be a *fait accompli*. But this of course is simply not true. Forming an intention to act is not like setting an alarm clock; it does not automatically cause action.

If on the other hand we accept some limitations on when or how intention can cause action (e.g., when the agent does not change his mind about the intention or is not prevented from acting), then we can accommodate cases of bare or future-directed intention. We may be able to accept that intention causes action provided certain other conditions are met. Unfortunately, this leads us into further trouble. Under this limited interpretation, to say that intention causes action is akin to saying that the cue ball causes an object ball to move in a certain direction. It is a technically accurate but deficient picture of the entire event, since for example it mentions neither the cue stick nor the player who controls it. While some observers (e.g., physicists studying the mechanics of the interaction) might be satisfied with such a rudimentary description of the event, it seems that some very important items have been omitted. Similarly, to say simply that intention causes action is incomplete. What is missing is the role of the agent as the one who links his preferences to the anticipated results by solidifying his plans via intention formation. As a result, a mere causal description of the relationship may obscure the agent's moral responsibility by removing his direct association with the action and its outcome.

In short, while the causation interpretation may be correct as far as it goes, it does not say enough about how intention and action are related. That more complete explanation must include not merely a description of the causal chain of events, but also an adequate account of both the agent and his preferences, an account which the linear relationship model provides.

4. OBJECTION TO INTENTION-ACTION INDEPENDENCE

Quite apart from the question of the suitability of the proposed model, it may be argued that intention is not a proper subject for independent moral evaluation, since it cannot be severed, conceptually or otherwise, from the action upon which it is directed. As Finnis & co. have suggested, intention is not separate from the act intended, but is 'the beginning of the act itself; the intention is seen as part of the action, with the same moral quality as the whole.'[20]

However, this conception of the relationship is fraught with difficulties. To begin with, it is unintelligible. Acts are not like objects; it makes no sense to talk as if they can be divided into constituent parts. So to describe intention as a part of action is to display a fundamental confusion about the nature of action itself.

But even assuming that such a partitioning of action is possible, the meaning of 'intention as part of action' is ambiguous. It might mean, for instance, that intention is an *integral* part of action, that is to say, action requires intention, and thus action without intention is impossible. However, this implies that all acts must have intentions, which leads to the rather absurd conclusion that actions which apparently have no (further) intentions (including both voluntary actions and the intentional actions which are done simply for their own sake) either are not acts at all, or else contain some 'hidden' intention.

There is however a second meaning of 'intention as part of action,' viz., that all intentions are *tied to* actions, that intention is, as it were, subsidiary to action. This in turn could mean either that there cannot be intention without action, or else that all intentions are 'actionable'. Under the first interpretation, intentions are incapable

---

[20]Ibid., p. 80.

of autonomous existence; intention without action is impossible. But this of course eliminates the possibility of any future-directed intention and means, for instance, that my 'intention' to see *Hamlet* becomes the genuine article when, and only when, I carry it out. So if my car breaks down on the way to Stratford, preventing me from seeing the play, we are forced to conclude that I never intended to see *Hamlet* at all. Surely, this is not right; an unfulfilled intention is still an intention.

Alternatively, to say that intention is tied to action may be simply to say that all intentions are 'actionable', viz., that an intention must be directed on some action. And this seems right, for an intention *is* an intention *to do* some action. Unfortunately, although we have finally uncovered an acceptable interpretation of the Finnis view, we have found ourselves further from our goal than when we started! For to say that all intentions are actionable is to say nothing more than all intentions tend toward states of being inclined to action, which is well short of even saying that intentions *lead to* action.

There is another, deeper problem with the Finnis view as it is integrated into their larger argument against nuclear deterrence. They make the claim that intention is part of action in order to bolster support for the wrongful intentions principle by showing that 'one's intention is morally more basic and *more important* than any performance or behaviour by which that intention is carried out.'[21] The choices made in intention formation are character-shaping: 'When one chooses a certain course of action [i.e., forms an intention], one determines oneself to be a certain kind of person.'[22] Thus it follows that 'intentions formed in the heart can be seriously wrong even if they are never carried out.'[23] And this of course is the essence of the wrongful

---

[21]Ibid. [emphasis added].

[22]Ibid.

[23]Ibid., p. 79.

intentions principle.

The main problem with this argument is that it fails to adequately account for two classes of intention, conditional intentions and unlikely intentions. We have already discussed conditional intentions in connection with the linear relationship model. With regard to the second, even though Finnis & co. briefly discuss unexecuted intentions,[24] they fail to come to terms with intentions which are not merely unfulfilled, but are unlikely ever to be carried out. When applied to these sorts of intentions, the blanket statement that intentions are character-shaping seems an oversimplification. Among other things, it leaves one with the distinct impression that agents who form, but do not execute, immoral intentions are in the same moral boat as those who carry out such intentions: 'Those who intend to perform wrongful acts and are prevented from doing so by circumstances beyond their control are considered blameworthy, *like those* who succeed in doing similar wrongful acts.'[25] Admittedly, there is something at least *prima facie* wrong with forming immoral intentions, but that alone does not justify equating (1) an agent who intends and does not act, with (2) an agent who acts. Neither can one equate (1a) an agent who forms an immoral intention and is then prevented from acting, despite his effort, with (1b) an agent who forms an immoral intention which is, *and is known by him to be,* extremely unlikely to be fulfilled. Perhaps this can best be seen by considering agents who form morally praiseworthy intentions.[26] Does an agent (1a) who intends personally to assist drought victims in Ethiopia but is killed enroute really deserve no more moral approbation than one (1b)

---

[24] Ibid., p. 83.

[25] Ibid., p. 80 [emphasis added]. Although Finnis & co. make an effort (pp. 81 n.9 and 99) to show that their position allows for a moral distinction to be drawn between intending alone and acting on that intention, it does not seem to adequately blunt the criticism that their argument in support of the wrongful intentions principle implies a blurring of this distinction.

[26] This assumes of course that if a wrongful intentions principle is justified, then a rightful intentions principle would also be acceptable.

who intends 'someday' to help those victims? Surely not.[27] Yet by the reasoning used to support the wrongful intentions principle, Finnis & co. would have us believe that there is no difference between the two.

This problem of unlikely intentions is especially damaging to them, since they are arguing against the justification of an agent who adopts a policy of nuclear deterrence. Such an agent is even further isolated from the act than is agent (1b), since not only does he consider his carrying out the retaliatory intention to be a very remote possibility, but he genuinely believes that the intention itself will very likely prevent him from having to execute it. There is thus something counter-intuitive about condemning him as just as bad as agent (2) or even as agent (1a). The Finnis view then is unsatisfactory and thus fails to damage the acceptability of the claim that intention is a proper subject for independent moral evaluation. On any interpretation which takes us beyond our starting point, we are forced into absurdity.

5. INTENTION AND NUCLEAR DETERRENCE

Admittedly, the above discussion only roughly lays out some of the more common sense notions inherent in the concept of intention. While it offers a plausible schema for understanding the relationships among preferences, intentions, acts and results, as well as the agent's role in forming those relationships, it remains silent on the precise nature of each of those entities. But it is the intuitive concepts in particular which highlight the important differences between intentions *simpliciter* and the sort of intention which forms an essential part of a successful nuclear deterrence policy. It is to that subject which we now turn our attention.

---

[27] Peter Abelard would disagree, claiming that 'each is as deserving as the other.' See Kenny, 1973, p. 137.

PART III: THE ROLE OF DETERRENT INTENTION

A cursory glance at any volume of the *Philosophers Index* over the past seven years will reveal that the issues of nuclear war and nuclear deterrence deeply concern not only strategists but moral philosophers as well. In addition to scores of articles in various anthologies, at least three journals have devoted entire editions to the moral problems posed by nuclear weapons.[1]

A comprehensive review of the literature related to this thesis would therefore expand it well beyond acceptable limits. As an alternative to such a review, I have selected for examination five critical thinkers on the more limited subject of intention in deterrence: Gerald Dworkin, Anthony Kenny, Kenneth Kemp, Gregory Kavka, and Michael Novak.[2] These philosophers have been singled out because, although they approach the subject from disparate angles, and arrive at often opposing conclusions, each brings to light important aspects of the concept of deterrent intention. And although none has fully and completely grasped that concept, their separate contributions constitute most of the relevant pieces of the puzzle of

[1] *Ethics* 95:3; *The Monist* 70:3; and *The Canadian Journal of Philosophy*, Supp. Vol. 12.

[2] Because of the subject matter of Chapter 3, I have chosen to critique the work of Finnis & co. at that earlier stage.

deterrence, a puzzle which I shall piece together in Part IV.

However, before beginning that examination, we shall consider whether the intention to retaliate is necessary for effective deterrence by investigating several different options in which that intention is absent, thus rendering deterrence a bluff.

4: CAN DETERRENCE BE BASED ON A BLUFF?

---

1. BLUFFING THEORIES OF DETERRENCE 52

2. THE STANDARD THEORIES 54

   2.1 The Inner Circle Bluff 54
   2.2 The Democratic Bluff 56
   2.3 The Overt 'Bluff' 57

3. THE ATOMISTIC BLUFF THEORY 61

4. THE BENEFITS OF EXAMINING BLUFFS 65

4: CAN DETERRENCE BE BASED ON A BLUFF?

A strange game. The only winning move is not to play.

--John Badham

As we have seen, the moral problems for nuclear deterrence concentrate around the intention to retaliate. Before examining a number of diverse views of that intention, it is worth asking if deterrence including the intention represents the only model for maintaining national security in the nuclear age. In other words, is forming the intention the only way to achieve that preferred goal?

At first blush, it seems the answer is no, for there are at least three other possibilities. The first of these would deny the necessity of making any threat at all. For example, the reliance on deterrent threats might be supplanted by a reliance on positive offers of mutual gain, which in the parlance of international diplomacy are referred to as Confidence Building Measures. Until recently it would have been utterly ineffective to attempt to replace deterrence with such positive offers, at least at the superpower level. However, events over the last two years (the 1987 Intermediate Nuclear Forces Treaty, moves to eliminate chemical weapons, and negotiations aimed at a major reduction in strategic arms, to name but a few), and the changes in attitude

which have precipitated those events, as well as the restructuring of Eastern Europe, have demonstrated that cooperative offers may at least obviate, and perhaps eliminate, the need for deterrent threats. While exploration of this alternative would take us far afield, we should note that this type of cooperation may eventually remove the danger of superpower confrontation, although the need for a credible threat may still remain in other areas of security maintenance, e.g., with other, less cooperative, countries.

The second possible alternative is to supplant the nuclear deterrent threat with a non-nuclear threat, eliminating the nuclear arsenal in favour of sophisticated conventional weaponry. While a thorough examination of this interesting possibility would also divert us from the main business of this thesis, I mention it only to point out that recent advances in weapons technology, especially in the area of targeting accuracy, make it likely that, for a counterforce target such as a hardened missile silo, a non-nuclear warhead may produce the same result as its nuclear counterpart, with significantly less collateral damage and a reduced danger of escalation.[3]

The third alternative would deny that a genuine intention to retaliate is necessary for effective deterrence. This would result in some form of bluffing deterrent to whose examination we now turn.

1. BLUFFING THEORIES OF DETERRENCE

In attempting to find a moral alternative to the accepted version of nuclear deterrence, many critics have considered, and rejected, a bluff theory of deterrence.

---

[3]For a discussion of the impact of emerging technology on the morality of deterrence, see Wohlstetter, 1983a, esp. pp. 22-23.

Such a policy would substitute an 'insincere threat'[4] for the real intention in an effort to avoid the negative implications of maintaining an immoral intention. Gregory Kavka, for example, looks to the possibility of a bluff after he admits that the self-corruption necessary for an effective deterrent (viz., the need for an otherwise moral agent to form the immoral intention to retaliate) 'is very likely to fail.'[5] This failure, coupled with the recognition by many strategic thinkers that the effectiveness of deterrence depends to a large extent (if not wholly) on the adversaries' *beliefs* about a deterrer's intentions, rather than the actual intentions themselves, leads naturally to the exploration of bluffing.

There are usually considered to be two main theories concerning deterrent bluffing. The first of these describes a scenario in which the deterring agent has not (yet) formed the intention to retaliate, despite having prepared to retaliate. While this strategy of 'keeping the option open' may not appear to be a classic bluffing strategy, it does contain a bluffing deception, since the real intention (to postpone the decision to retaliate) differs from the apparent intention inherent in the preparation, viz., to retaliate if attacked.[6]

The other bluffing theory, and the one which commands more attention, is a standard overt deterrence policy accompanied by the covert decision never to form the intention to retaliate, even if attacked.[7] This theory itself can appear in one of two forms. The first is what I call the Inner Circle Bluff, in which the covert lack of intention is kept a closely guarded secret among those very few national leaders ultimately responsible for ordering retaliation. All others, those in the execution chain

---

[4]Morris, p. 481.

[5]Kavka, 1978, p. 296.

[6]Finnis, et al., p. 114.

[7]This is the version which, e.g., Kemp (1987a, p. 277) and Dworkin (1985, p. 448) examine.

of command, ordinary citizens and (of course) the enemy, are unaware of the absence of a genuine intention to retaliate. The second form may be called the Democratic Bluff, where the secret is shared by *all* those in the execution chain, although remains unknown to ordinary citizens and the enemy.

In addition to these, there are two other versions of the bluff: Anthony Kenny's Overt 'Bluff', where the absence of intention is publicly announced, and what may be called the Atomistic Bluff, where that absence is unknown to anyone except the agent himself. Of these last two, Kenny's Overt Bluff shares an important feature with the standard theories in that it could possibly be implemented and enforced as national policy. The Atomistic version cannot.

2. THE STANDARD THEORIES

Any form of bluff would seem to be morally superior to a policy which includes the murderous intention to retaliate. Indeed, critics of bluffing rarely attack it on moral grounds, save to raise the relatively minor point (considering the enormous evils of non-bluffing strategies) that bluffing requires dissimilation or deception, both of which are at least *prima facie* wrong. By far the more potent objections are practical rather than moral, where the main emphasis is on the claim that bluffing deterrents are ineffective.

2.1 The Inner Circle Bluff

The primary assumption in the Inner Circle Bluff is that the secret must be kept closely guarded and limited to a very few individuals. However, considering that deterrence is not meant to be a short term enterprise, but must be transferred (at least

in the case of western democracies) to succeeding administrations or governments with perhaps widely divergent world views, the theory is 'highly implausible.'[8] Furthermore, such deception not only may be impossible in an open society, but may also be inconsistent with some of the basic values of such a society, e.g., democratic participation in government, freedom of information, etc.[9] And even if the secret could be kept, the Inner Circle Bluff would still require those outside the circle, but yet vital to the success of the deterrent machinery, to maintain the real intention to retaliate.[10] And those inside, as directors of the deterrent mechanism, would continue to bear at least indirect responsibility for that intention.

There is an additional practical problem with keeping the secret. As Finnis & co. report in their succinct survey of western deterrent force capabilities, in the event of a 'decapitation' strike which eliminates a nation's major command and control centres, the nuclear deterrent network is designed to revert to a 'fail-deadly, rather than fail-safe mode,' thereby ensuring retaliatory launch in the absence of a direct countermanding order.[11] Under the Inner Circle Bluff theory, the decision to eschew the intention will die with the members of that circle, and result in retaliation despite their efforts. Thus a morally upright group of leaders would, at best, be gambling on preventing retaliation.

---

[8] Finnis, et al., p. 116.

[9] Morris, p. 481.

[10] This point is made by both Kemp (1987a, p. 277) and McMahan (p. 524). Against this, Dworkin (p. 454) argues that the outer circle agents need not intend to *retaliate*, but merely intend to *obey orders*. But this of course does not get them off the moral hook, unless members of the military and others in this group are automatons (as Dworkin thinks they are) who blindly and irrationally follow the orders of their superiors, a quality which Dworkin, if he thinks through the implications, must surely agree is not a desideratum.

[11] Finnis, et al., pp. 56-58. See also Dummett, p. 120; and Hardin, pp. 171-72.

2.2 The Democratic Bluff

These and other problems with the Inner Circle Bluff lead naturally to the second theory, in which all those in what Jeffrie Murphy calls the 'chain of agency'[12] (i.e., those responsible for executing retaliation) are let in on the secret, and also decide not to form the intention carry out the deterrent threat.

Although it shares some of the same problems as the above theory (e.g., the general public would still maintain support for the murderous intention), this scenario does solve some of the earlier nagging difficulties. For instance, a decapitation strike would no longer result in retaliation, nor would those at the top be responsible for fostering the immoral intention on the part of those whose job it would be to execute the orders.

However, the Democratic Bluff falls victim to several new and perplexing difficulties. It would, for instance, be much more difficult to keep the secret from the enemy, a fact which might undermine the effectiveness of the deterrent by inviting the enemy to test the deterring nation's resolve.[13]

A number of different critics also discuss one final but devastating 'institutional' problem with the Democratic Bluff (and really all bluff theories): Deterrence is a *social* undertaking which requires the consent and trust of the governed. This idea impacts bluff theories in two different ways. The first, identified by Dummett, is that acceptance of the possibility of bluff is an irrational act of faith by the electorate. In a flurry of rhetorical wit if not sound argumentation, he attacks the rationality of such acceptance:

> This faith is utterly blind; everything tells against it. . . . What sense does it make

---

[12] Murphy, p. 531.

[13] See Kemp, 1987a, p. 278; McMahan, p. 524. It should be noted that other writers dispute this claim, based on the argument that the effectiveness of deterrence is a function of the enemies beliefs about the likelihood of retaliation. See references in note 3, Chapter 2.

> to trust politicians--any politicians--not to do what they say they will do? Politicians, in power or out of it, lie as a matter of course, a fact to which there are countless attestations. They cannot be trusted to do what they say they *will* do; how can they be trusted in this instance *not* to do what they say they will do?[14]

Finnis & co. also attack bluffing on institutional grounds. It is not simply that a bluff won't work, but rather that it is analytically impossible for a nation *qua* nation to bluff about deterrence, just as it is impossible for a team *qua* team to play to lose, despite the actions of its individual players. The purpose of a team is to play to win, even if all individual members conspire to throw the game. For the actions of the team as a unit can only be properly understood 'as contributions to the social act of the team: playing to win.'[15] In the same way that a team's actions constitute a unified whole, so also a deterrent system displays the unity of a single social act. And that choice of a community *qua* community *cannot* be a bluff.

2.3 The Overt 'Bluff'

The last of the standard theories has been suggested by Anthony Kenny as part of his proposal for disarmament.[16] It is standard in that it can be implemented as a national policy, but it is nonstandard in that it requires no deception. The basic idea is that, during the disarmament process, deterring nations should form *and announce* a real and credible intention never to launch their nuclear weapons. That is, they should make public what was secret in previous bluff scenarios. Each nation then continues to disarm but retains a small cache of weapons as 'some guarantee against

---

[14]Dummett, p. 121.

[15]Finnis et al., p. 121. Although I believe this analogy and therefore the objection raised to be faulty for several reasons (e.g., it is not entirely clear that a team cannot play without the corporate purpose of winning the game), I shall forego a more formal critique in favour of highlighting an alternative bluff theory, which will perhaps better serve to disprove the Finnis claim that deterrence cannot be based on a bluff.

[16]Kenny, 1985, pp. 70-101. James Sterba (p. 101) presents a similar, but less convincing, argument for the theory.

bad faith' by its adversaries.[17] Such an arsenal would still constitute an effective deterrent, given the fact, as we have seen, that effectiveness is a matter of one's enemies' perceptions and beliefs.

We should underscore the fact that since this proposal involves no attempt at deception, it is not a bluff in the usual sense of the term, except perhaps that it may be considered a type of 'double bluff' where the nation who possesses the nuclear weapons reaps the benefits of its adversary's doubts.[18] But neither is it a straightforward policy, since the existential deterrent is maintained without its usual accompanying intention.

The proposal has obvious strategic advantages over other standard bluffs, the effectiveness of which is undermined by the possibility of a security leak. In Kenny's version, there is no danger of a leak because there is nothing hidden. It also has significant moral advantages over deterrence with the intention to retaliate since

> Continuing to maintain the physical operability of the nuclear weapons with the sole purpose of using them as bargaining counters to secure balanced and eventually total reduction of Soviet forces would not involve even a conditional willingness to use the weapons in any warlike role.[19]

The deterrent value of the remaining weapons (as one winds down toward total disarmament) would be an unintended beneficial side effect.

There are several important objections which can be raised against the proposal, and Finnis & co. spend some time (albeit in an endnote) detailing these.[20] First, despite the renunciation of use, the beneficial side effect still arises from an implicit threat

---

[17]Ibid., p. 98.

[18]Fisher, p. 75.

[19]Kenny, 1985, p. 98.

[20]Finnis, et al. pp. 125-27.

to commit a prohibited violent act. Retention of nuclear weapons is *prima facie* inconsistent with a genuine intention never to use them. This inconsistency can only be resolved in terms of a double bluff, where the deterrer seeks to gain ground without sacrificing his own morality. But, Kenny would respond, to claim that there is such an implicit threat in this proposal is to misunderstand the true nature and depth of the renunciation of use. Not only is there no intention, there is also neither threat nor willingness, but simply manifest ability. That alone cannot be immoral, even if it results in benefits which would otherwise have been achieved immorally.[21]

Then it cannot be effective as a deterrent, or even as a hedge against bad faith. For it is impossible and incoherent for soldiers to train to do that which they are ordered never to do. The proposal is thus either a disguised version of bluff, or else a 'pointless drill' for those practising to carry it out.[22]

Two responses can be made here. First, today's soldiers carry on training for retaliation with the 'profound hope' that they will never have to demonstrate the effectiveness of their training; this proposal merely solidifies that hope.[23] Moreover, today's soldiers regularly (and often unreflectively) conduct training exercises with no anticipation of actually using their acquired skills against an enemy. So if having a point means training for actual utilisation, then the current drills are pointless as well. And when soldiers do pause to reflect on their role in deterrence, it is not simply with a 'profound hope', but with a real belief that their efforts are actually preventing the need to exercise their abilities. This attitude would certainly continue if the proposal were to be adopted as policy.

Secondly, to see the retention of some weapons as pointless is to assume that

---

[21]Kenny, 1985, pp. 98-99.

[22]Finnis, et al., p. 126.

[23]Kenny, 1985, p. 80.

the only 'use' for nuclear weapons is 'launch and detonation.' But there can be other uses. Indeed, the last four decades are evidence that a nuclear arsenal can be used, without detonation, as a deterrent to aggression, a role which should not be lightly dismissed. And here Kenny is simply proposing another (non-explosive) use, i.e., as a bargaining chip to ensure trustworthiness.

The last Finnis objection to the proposal is that the elaborate and prolonged debate preceding such a decision for policy change in a democracy would undermine the effectiveness of the bargaining chip.[24] However, this objection is merely a disguised version of the standard argument against the bluff theory, viz., that it cannot be kept a secret in a democracy. Kenny's proposal avoids that objection since the renunciation of use (i.e., launch and detonation) is, both during and after the debate and decision, overtly announced. This is not a case of insincere threat because no threat is made. Thus it is not really a bluff, since, so to speak, all the cards are on the table from the start. The retention of a small nuclear force as a bargaining chip and hedge against bad faith may indeed have the (foreseen) side effect of instilling fear and thus preventing cheating. However the purpose, the aim, is not to frighten or threaten, but to maintain a credible position during the process of disarmament. The Kenny proposal may suffer from other difficulties (e.g., the impossibility of complete nuclear disarmament, since the technology cannot be dis-invented[25]), but Finnis & co. have not uncovered one here.

---

[24]Finnis, p. 126.

[25]Fisher (p. 67) says of this problem: 'The risk would still remain that under the threat of imminent or actual war, the nuclear knowledge--which can never be erased from human consciousness--would be reapplied.'

3. THE ATOMISTIC BLUFF THEORY

The standard bluffing theories are fraught with problems, and thus rightfully condemned by both opponents and supporters of deterrence. There is however a completely nonstandard version which we should consider before abandoning the idea altogether.[26] This theory may be called the Atomistic Bluff because it lacks the 'corporate' assumption which can be found in all other standard bluff theories, and which is the source of all ineffectiveness objections. That assumption is that any bluff must be a corporate enterprise, that is, all of the parties privy to the strategy must be aware of the bluff, and *all must be aware of the corporate agreement to the strategy.* It is this last requirement, that each of the players knows the true intention of the others involved, which makes the bluff unworkable since, it is claimed, the secret cannot be kept for long.

But the Atomistic Bluff theory lacks this damaging assumption. In this scenario, there would still be public assent to the retaliatory intention, but at the same time at least some member(s) of the society would maintain a private and *uncommunicated* refusal either to form the intention or to carry out the intended retaliation. This private denial of the deterrent threat could either be held by the person at the apex of the execution pyramid, or alternatively be held universally throughout the society. In either case, the theoretical society would have the external appearance of any of the current nuclear superpowers with regard to the resolve to retaliate. What it would lack is the genuine, but unobservable, intention to act.

The first objection which can be leveled at the theory is a relatively minor complaint that it is nonfalsifiable, and thus immune from counter-argument. But while this may be true, it is not unique to the Atomistic Bluff. For this quality is shared with

---

[26] I am grateful to Jonathan Glover for suggesting this version to me.

all bluff theories,[27] and indeed with all questions about intention, a mental phenomenon which Anscombe rightly notes is 'purely interior.'[28] Thus it is certainly not a criticism of the theory alone, and may not be a sound criticism at all.

Similarly, it may be objected that the theory suggests a maxim for action which must necessarily fail the test of publicity, a constraint which John Rawls has stated must be satisfied by any acceptable moral principle, and which Sissela Bok has set forth as a criterion for the moral justification of action. The idea that each participant in nuclear deterrence, in order to rescue his or her morality, must maintain and conceal an intention to thwart the execution of the deterrent violates this publicity constraint. As Bok puts it, 'A secret moral principle, or one that could only be disclosed to a sect or guild, could not satisfy such a condition.'[29] This additional constraint cannot be subsumed under the criterion of universality, since 'it is possible that all should understand and follow a principle and yet this fact be not widely known or explicitly recognized.'[30] The two constitute independent criteria. Universality requires that a moral principle be applied to every relevant moral agent; publicity demands that adherence to the principle be publicly acknowledged. So although it could be argued that the maxim generated by the atomistic bluff theory satisfies the condition of universality (it applies to everyone in the execution chain), it cannot be publicly acclaimed without suffering from the efficacy problems of corporate bluff theories.

Two answers may be formulated against this objection. First, the maxim envisioned cannot and should not be considered to be a general moral principle of the type which Rawls and Bok have in mind. It is merely a suggestion of a possible

---

[27] Finnis, et al., p. 115.

[28] Anscombe, 1966, §4.

[29] Bok, p. 92.

[30] Rawls, p. 133.

state of affairs in which all participants in deterrence secretly lack the intention to retaliate. Therefore the publicity constraint, much like the requirement for falsifiability, cannot be effectively used against it. Secondly, even if this were a legitimate complaint against the theory, it is no more potent than would be a generic objection against all forms of deception, which of course would apply to all bluffing theories. For, given the enormous potential danger associated with the alternative, viz., a deterrence policy which includes a genuine intention to retaliate along with its increased risk of execution, and barring a defensible absolute prohibition against deception, this complaint seems to pose a relatively minor problem for the theory.

More so than the theoretical problems of falsifiability and publicity, the atomistic bluff theory is open to the attack that it is extremely implausible, depending as it does on an unwarranted assumption about the nature of all individuals associated with deterrence. This is made clear by the fact that it cannot be enforced as part of a national policy. It can at best be accepted on faith alone. And so once again Dummett's words apply with damning effect: 'This faith is utterly blind.' But while this may be so, perhaps the theory is not as far-fetched as it first appears. Consider President Reagan's thoughts on the alternative:

> Think of it. You're sitting at that desk [in the Oval Office]. The word comes in that they [the missiles] are on their way. And you sit there knowing that there is no way, at present, of stopping them. So they're going to blow up how much of this country we can only guess at, and your only response can be to push the button before they get here so that even though you're all going to die, they're going to die too.... There's something so immoral about it.[31]

Admittedly, Reagan spoke these words in the midst of his campaign for support for the Strategic Defence Initiative, but one can still read the message, only just below the surface, that he, surely like all other leaders of nuclear powers, has grave doubts

---

[31] Sidey, 1985, p. 29.

about whether he would order the planned retaliation.[32] Surely anyone, including those whose job it is to carry out the retaliation order, who takes a moment to consider the overwhelming absence of reasons in favour of retaliation after deterrence has failed must begin to doubt the rationality of such a move. And this conclusion, coupled with the belief that keeping the deterrent credible is the very thing which prevents one from being forced into that irrational corner, could well lead each of those individuals in the execution chain of command to (secretly) adopt the intention not to carry out retaliation. Nor, for that matter, must *everyone* be bluffing in this way. For any attempt to retaliate would certainly be thwarted even if only some critically placed individuals were to adopt an atomistic bluff.

But the question remains, does this theory survive Dummett's attack on bluffing? No. But the attack is irrelevant to the morality of the policy. We can distinguish two questions here. (1) Can a citizen be certain that his nation's leadership will act in accordance with an atomistic bluff policy? (2) Would it be morally acceptable for a deterring agent to adopt a policy of deterrence suggested by the atomistic bluff theory? The answer to (1), the question at the heart of Dummett's attack, is obviously no. But the question of moral relevance is (2), and the answer to that may well be yes. And even with respect to (1), the theory does take a step in the right direction. It shows that a citizen may conclude that there is at least some possibility, however slight, that his country's deterrence policy does not include a murderous intention to commit pointless genocide. For if he has concluded that retaliation is pointless and immoral, then there is a chance that his fellow citizens, some of whom are responsible for the execution of retaliation, have arrived at the same conclusion. Thus he might reasonably believe that the actual chance of retaliation occurring has been reduced

---

[32] Finnis & co. (pp. 117 and 130) have uncovered further evidence to support the claim that at least western leaders are equally concerned about the immorality and indeed the irrationality of retaliation. The authors believe, however, that the evidence cannot support any bluff hypotheses.

because of the absence of the intention to do so.

One final objection must be answered, viz., that the 'institutional' problem which plagues other forms of bluff must also damage the credibility of this atomistic version. For if deterrence is a social act, then it is not individual private intentions, but the collective public intention of the nation which is the critical determinant of moral acceptability. And that public intention remains under any bluff scenario. But deterrence as a social endeavour is condemned by these critics not simply *qua* endeavour, but because it relies on the real intention to do wrong. The Atomistic Bluff does not rely on that intention. All that the theory claims is that each individual *qua* individual, not *qua* member of the deterrent force, secretly lacks the intention. This is consistent with outwardly participating in deterrence *up to the moment of execution*, when of course one would no longer be participating in deterrence at all. At that critical moment, the response to attack would look radically different from the apparent deterrent policy to that point.

4. THE BENEFITS OF EXAMINING BLUFFS

What is the point of arguing for a largely nonfalsifiable theory about deterrence? Realistically, appeal to the possibility of bluffing cannot save deterrence since the mere hope of deception is, in David Fisher's words, 'a base too fragile on which to rest our security for all the, perhaps lengthy, time that deterrence may be required.'[33] Indeed, it is a base too fragile for our morality as well. Thus, while it (obviously) does not wholly rescue the policy from the flames of damnation, the theory does begin to cast doubt on the seemingly well-laid arguments which purport to condemn deterrence. For that condemnation results from the perceived necessity that deterrence must include the

[33]Fisher, p. 117.

intention to retaliate. But if there is no necessary connection between the two, perhaps nuclear deterrence is not as morally deficient as its critics believe.

5: OTHER VIEWS OF DETERRENT INTENTION

1. DWORKIN: EMBEDDED INTENTIONS 68

   1.1 The Peculiarity of Deterrent Intention 68
   1.2 The Morality of Deterrent Intention 71

2. KENNY: INTENTION AND USE 73

   2.1 Structure of the Argument 73
   2.2 The Threat Dissected 75
   2.3 The Deterrer's Intentions 79
   2.4 Kenny's Contributions 81

3. KEMP: INAPPLICABILITY OF THE WRONGFUL INTENTIONS PRINCIPLE 82

   3.1 Must Deterrence Threaten Innocents? 83
   3.2 The Special Nature of Deterrent Intention 84

4. KAVKA: THE PARADOXES OF DETERRENT INTENTION 86

   4.1 The Wrongful Intentions Paradox 87
   4.2 Rejection of the Bridge Principles 92

5. NOVAK: DISSECTING DETERRENT INTENTION 93

   5.1 The Several Species of Deterrent Intention 95

5: OTHER VIEWS OF DETERRENT INTENTION

A large number of philosophers have focused their studies of deterrence on the intention involved in the policy. I have chosen for examination five of these studies, in which the authors seem to have made significant progress towards unraveling the exact nature and role of deterrent intention. We shall examine them in turn.

1. DWORKIN: EMBEDDED INTENTIONS

In a thoughtful and balanced paper, Gerald Dworkin examines the problem of deterrent intention and its relationship to actual use from both consequentialist and deontological perspectives. He concludes that neither use nor the genuine threat of use of nuclear weapons (which must include the intention) is permissible.[1]

1.1 The Peculiarity of Deterrent Intention

Dworkin begins his discussion of intention by arguing that intention is indeed

[1] Dworkin, pp. 445-60. Interestingly, he believes that the only morally acceptable deterrent would be one which did not include the intention to retaliate, i.e., a bluffing deterrent.

a proper subject for moral evaluation. While this claim is certainly acceptable (we have seen in Chapter 3 that this is so), only one of his several reasons in support of the claim--that intentions as mental acts are subject to the same scrutiny as ordinary acts--produces an acceptable argument. The others are either too elliptical to stand without further support or else simply fallacious.[2] Despite these problems, Dworkin rightly recognises the importance of first establishing the independence of intention before examining its role in deterrence.

After setting out the relevant features of intention, Dworkin discusses four peculiar aspects of deterrent intention which set it apart from ordinary, and even conditional, intentions. Of these, he seems to think that only two are genuinely significant. Indeed, the first two, self-frustration and the production of autonomous effects, he mentions and dismisses in a single passage: Forming the conditional intention to retaliate, he argues, has two causal consequences: it increases the likelihood of the act occurring and it produces autonomous effects. Since the act of retaliation is ruled out on consequentialist grounds, any action which increases the likelihood of retaliation occurring is also ruled out, even though it appears that the production of autonomous effects provides evidence in favour of forming the intention. And since forming the intention does increase such likelihood, it must be ruled out, regardless of any other considerations, including the fact that the intention is formed (at least in part) to prevent the occurrence of the conditions which would lead to its execution:

> Since a consequentialist theory is concerned with the goodness or badness of states of affairs, the relevance of the forming of an intention is exhausted by its causal contribution to the production of one or another state of affairs. . . . In short, the relevance of intentions to do morally forbidden acts is exhausted by the increased risk of harmful consequences.[3]

---

[2] He says, for example, that an intention can be wrong 'because of the kind of intention it is.' (p. 447). This of course begs the question about the possibility of moral evaluation of intentions.

[3] Ibid., p. 450.

Deterrent intention increases the likelihood of the wrong act being performed, therefore its formation is wrong.

But there is a problem with Dworkin's analysis. In it, he only considers one possible state of affairs, i.e., the occurrence of the wrong act (and its effects). It is this state of affairs which is ruled out on consequentialist grounds. However, as Dworkin admits, this is only 'normally' the case. While this seems acceptable for ordinary intentions, the existence of the autonomous effects produced by deterrent intention (i.e., prevention of the relevant conditions) highlights another relevant state of affairs which includes the effects achieved by the formation of the intention to act (without the act itself being performed). The formation of the intention, while it admittedly increases the risk of execution, also makes a causal contribution to the production of this other state. Thus the importance of the autonomous effects cannot be dismissed.[4]

The two oddities of deterrent intention to which Dworkin does attach at least some importance are the fact that the act of retaliation is not valued by the deterring agent, either as a means or an end, and the fact that the agent believes that forming the intention *decreases* the likelihood of execution. As we mentioned above and in Chapter 3, an agent forms an intention in order, among other things, to increase the likelihood of performing the intended act and thereby to realise some desirable state of affairs. Not so with the deterring agent. He believes that his intention formation will render it less likely that he will act as he intends. Thus he cannot value the act while at the same time working to prevent its occurrence. His belief about the effect of intending is crucial: 'Since this is the point of forming the intention, it is part of the logic of deterrent intentions that one does not have to value the fulfillment of the intention, either as an end in itself or as a means to some other end one has.'[5]

---

[4] Indeed, we shall return to the subject in Chapter 6.

[5] Ibid., p. 456.

Dworkin's point is critical, but his context reveals an underlying misconception about the nature of deterrence, a misconception which causes him to miss the significance of what he says. For he immediately moves on to emphasise the retaliatory aspect of deterrence: The intention 'reflects the agent's values by showing what he is prepared to do *(under certain circumstances).*'[6] Surely the intention must also, and to a greater extent, reflect the agent's values by showing what he (at least believes he) is trying to accomplish by forming the intention. In addition, Dworkin emphasises that the agent (merely) believes that his intention will have the desired effect, implying that his belief may have no foundation in reality. But belief, i.e., perception, is central to deterrence effectiveness. Perception constitutes reality in the relevant political arena. If both parties in a mutual deterrence situation *believe* that forming the intention decreases the likelihood, then it does, since that belief is what serves to prevent either party from overstepping the accepted limits.[7]

1.2 The Morality of Deterrent Intention

Having spent the majority of his essay exploring the rationality of deterrent intention, Dworkin concludes with a discussion of the analogous moral question: 'Can it be moral to commit oneself to actions which, independent of the policy in which they are embedded, are immoral?'[8] The answer, he says, is no. Cases of paternalism, 'where we impose a risk of bringing about (otherwise impermissible) harm as the best chance of avoiding worse harm,' require that the risk be justified to the threatened

---

[6] Ibid. (emphasis added).

[7] For various views on the role of belief in deterrence, see, e.g., Fisher, esp. p. 79; Kavka, p. 287; Kenny, 1985, p. 79; Morris, p. 481; Schelling, p. 36; and Sterba, p. 101.

[8] Dworkin, p. 457.

party.[9] The justification of the threat in turn justifies the action itself. 'The direction of moral justification is from the conditional intention to the carrying out of the intention.'[10]

Where the persons at risk are precisely those whose unjust actions one is trying to deter, the justification is straightforward, given of course some proportionality limits.[11] But there can be no such justification for threats against the innocent. Therefore, deterrence is wrong because 'we aim at the death of particular persons as the means of securing whatever benefits are at stake.'[12]

We are not yet to the point where we may satisfactorily examine this last claim (except perhaps to point out that Dworkin errs in stating the aim of nuclear deterrence; the agent aims not at death, but only at the threat of death). But what we gain from Dworkin are his three primary contributions to the study of deterrent intention. The first of these, underlying his above discussion of the morality of the intention, is embeddedness: The intention to retaliate is formed within a larger context within which it must be evaluated. Even though Dworkin finds that deterrent intention lacks the necessary justification, his recognition of the existence of a morally important context is a significant development in the effort to uncover the true nature of deterrent intention.

His second contribution is to highlight the deterring agent's belief about his intention. There is moral significance in the fact that the agent believes that maintaining his intention diminishes the probability that he will have to act on that intention. As we shall see in Chapter 6, it is this feature of deterrence which leads

---

[9] Ibid., p. 458.

[10] Ibid., p. 459. This is the argument David Gauthier (esp. p. 483) employs to support his claim that retaliation is morally acceptable.

[11] Dworkin (p.459) mentions three possible justifications for the threat derived from the right to self-defence: the potential aggressor's forfeiture of his right to non-maleficence, the notion of 'fault forfeits first,' and the utilitarian concept of mutual gain resulting from permitting self-defence.

[12] Ibid., p. 460.

many critics to argue that deterrent intention is self-frustrating.

Finally, Dworkin recognises a distinct peculiarity about deterrent intention. This recognition is a first step towards realising that deterrent intention is inherently unique, even among conditional intentions. The nature and implications of this uniqueness are the subjects of Chapters 6 and 7.

2. KENNY: INTENTION AND USE

In *The Logic of Deterrence*, Anthony Kenny examines the most popular arguments for and against nuclear deterrence. Finding merits and problems with both sides, he argues for a *via media* between unilateralists and supporters of current deterrence policies, a thought-provoking proposal for a 'minimum transitional existential deterrent' as a step toward complete multilateral disarmament.[13] Although Kenny seems to miss a vital point about the actual effect of the retaliation intention on deterrence, it is worth spending some time analysing his arguments as a way of examining his comments on deterrent intention.

2.1 Structure of the Argument

Having argued in the first part of his book that there is no legitimate use for nuclear weapons, Kenny turns his attention to considering justifications of nuclear deterrence alone (i.e., without use). To be justified, such a policy would at least have to provide one's potential adversary with a reason to 'desist from action.'[14] But Kenny sees a paradox here: Either the threat to retaliate is genuine (viz., it includes an intention to retaliate), or else it is a bluff. But neither of these possibilities will deter

---

[13] Kenny, 1985, pp. 70-73.

[14] Ibid., p. 38. Kenny lays out the same argument in an earlier paper (1984, p. 20).

the enemy. For if the threat is real, then (for the reasons against actual use above) the enemy will believe the deterrer to be mad (irrational); if the threat is actually a bluff, then the enemy will see the deterrer as lying. In either case, the enemy will have no reason to be deterred.

But the enemy is deterred, as Kenny admits. And the reason is this: A deterrent posture is meant to 'provide an input to the practical reasoning of a potential adversary.'[15] Such an input does not necessarily have to be rational in order to have a bearing on that reasoning. Deterrence works because the reaction of the deterrer is (at least) *unknown*. An enemy cannot be sure of a purely rational response to aggression. As Kenny acknowledges, 'It is a nation's power, rather than its willingness, to use nuclear weapons, which is the essence of deterrence.'[16]

However, the real question engaging Kenny, and virtually all moral philosophers, is not whether deterrence is effective, but whether it is morally acceptable.[17] If it is really the power which deters, then it seems that the mere possession of a nuclear arsenal would be sufficient. For it is this 'existential deterrent,' coupled with the doubt in the mind of the enemy, which carries the successful deterrent message.[18] If so, then there is no need to add to that existential deterrent any declaratory policy, either genuine (viz., including a real conditional intention to retaliate) or phony (i.e., bluff).[19]

---

[15] Ibid., p. 47.

[16] Ibid., p. 53.

[17] Admittedly, some strict consequentialists might hold that the effectiveness of the deterrent answers the question about its morality. The only remaining question would then be an empirical one, viz., is deterrence in fact effective?

[18] Original use of the term, 'existential deterrent', is credited to McGeorge Bundy: 'My aim in using this fancy adjective is to distinguish this kind of deterrence from the kind that is based on strategic theories or declaratory policies or even international commitments. As long as we assume that each side has a very large number of thermonuclear weapons which could be used against the opponent, even after the strongest possible preemptive attack, existential deterrence is strong. It rests on uncertainty about what could happen, not in what has been asserted.' Quoted in Kenny, 1985, p. 51.

[19] Finnis & co. (pp. 105-10) argue that mere possession itself constitutes a real threat (viz., including the intention to use), and so is no more morally acceptable than a genuine stated threat: A deterrent system 'deters by constituting threats, and thus the very acts of obtaining and possessing the deterrent are *acts of making these threats*.' (p. 107).

However, Kenny believes that a nation cannot merely possess an effective deterrent. That power necessarily represents a willingness on the part of the individuals of that nation to exercise the deterrent force. That willingness is just as morally damaging as a genuine intention.

But it seems that Kenny is incorrect in equating power and willingness. The common notion of power is that of an ability or capacity to perform some function. We thus speak of, for example, the *power* of persuasion. The key here is the emphasis on potential which has not been actualized. 'Power' seems to be the equivalent of Hobbes' 'strength', where in speaking of individual human equality, he writes that 'the weakest has strength enough to kill the strongest.'[20] This usage does not seem to include any notion of willingness. It is certainly true that I have the power, in the intuitive sense of the term, to perform many repugnant acts which I have no willingness whatsoever to perform; similarly, there are many deeds I am willing, but nevertheless unable, to perform. It seems then that power and willingness are discreet notions, not necessarily connected in any way, except perhaps that both are essential elements in our intuitive understanding of intention. Kenny seems to have confused willingness with preparedness. For it is certainly true that, to be effective, a deterrent force must be prepared for use. But there are surely examples where one is prepared (i.e., trained), but not (yet) willing. One can be an expert marksman, yet not be willing to shoot anyone. Accepting Kenny's equation would lead to an errant definition of 'prepared' not merely as 'trained', but as 'trained and willing'.

2.2 The Threat Dissected

Kenny moves on to examine the threat (and thus the intention) to retaliate. He develops an argument against its formation which is based on his earlier claim that

---

[20] Hobbes, 1651, (Raphael, para. 47).

the actual use of nuclear weapons would be irrational. Since use is irrational, then intention to use is likewise irrational. Irrationality is not justified, so the intention to retaliate is not justified.

However, Kenny's earlier remark can be applied to his irrationality argument: It 'moves a little too fast.'[21] It is often unclear whether he is seeking to prove that the threat to retaliate is immoral or irrational, or both. He seems to be offering two intertwined claims, that the threat *cannot* (rationally) be maintained, and that the threat *should not* be maintained. The argument for the first of these claims seem to run something like this:

(1) If the threat to retaliate includes the intention to retaliate, then it is irrational.

(2) The threat does include the intention to retaliate.

(3) Therefore, the threat is irrational.

(4) One cannot perform irrational acts.

(5) Therefore, the threat cannot be maintained.

Statements (1), (2), and (4) require further support. Premise (1) is the conclusion of the argument based on Kenny's claim, discussed above, that actual use is irrational, and the assumption that since use is irrational, intention to use is also irrational. Unfortunately, this crucial assumption is not supported. Presumably Kenny believes that the truth of the implicit major premise--it is irrational to intend that which is irrational to do--is as self-evident as the wrongful intentions principle. But it is not all that clear that the wrongful intentions principle is self-evidently true. And even if it were intuitively acceptable, it is hard to see how that would impact on the irrationality premise in question. A prescriptive claim about the moral connection between act and intention cannot simply be rewritten into a more descriptive claim about the rational connection between the two. Neither is independent proof of the

---

[21] Ibid.

claim readily apparent. Assuming an intuitive sense of 'irrational', i.e., 'without good reason', it is not difficult to imagine a case (e.g., the intention to retaliate in nuclear deterrence) where an agent has reason to form an intention, but no reason to act on that intention. In this case the intention is rational, but the act is not, which leaves the truth of Kenny's assumption in doubt.

The claim of statement (2), that the threat to retaliate includes a genuine intention to do so, arises from Kenny's earlier argument that a (standard) bluff is not feasible.[22] And since the threat must be either a bluff or real (with a genuine intention to retaliate), we are left with the conclusion that the threat includes the intention. But as we have seen in Chapter 4, this chain of reasoning is suspect, since Kenny only considers 'corporate' bluff policies. Lack of awareness of the (admittedly remote) atomistic bluff theory leads Kenny and many other critics to reject categorically the possibility of bluff.

Finally, premise (4), that one cannot perform irrational acts, stands in need of some support. Presumably it is not merely a descriptive comment about the human condition, on the same order as 'one cannot live forever.' A glance at the evening newspaper will verify that not only is irrational action empirically possible, but that it is an all too frequent occurrence. Perhaps what is meant here is not that irrational acts are impossible, but that they are not (rationally) justifiable; what (4) really means is that 'one is not justified in performing irrational acts,' since they are performed without a justification-providing reason. But then the premise so interpreted becomes circular, viz., 'one is not justified in performing acts without justification.' The most acceptable reading of (4) is a prescriptive one which unfortunately slurs the distinction between Kenny's two arguments against the threat: 'It is wrong to perform irrational acts.' Read in this light, the premise is clearer (although still not self-evidently true). But now the conclusion (5) cannot be about the irrationality of the threat; it must

---

[22] Kenny, 1984, pp. 23-24.

instead be a claim about its immorality.

And here may lie the power in Kenny's argument. The claim is not about rationality *per se*. It is about the immorality of the threat to retaliate. If so, we may mirror the above argument to reflect that emphasis:

(1*) If the threat to retaliate includes the intention to retaliate, then it is immoral.

(2*) The threat does include the intention to retaliate.

(3*) Therefore, the threat is immoral.

(4*) One cannot perform immoral acts.

(5*) Therefore, the threat cannot be maintained.

As above, (1*), (2*) and (4*) stand in need of justification. Since (2*) is identical to (2), the same criticism applies to both, viz., that Kenny has failed to consider alternative bluff hypotheses. Premise (1*) is based on a modification of Kenny's argument about actual use, plus a direct appeal to the wrongful intentions principle. The modified argument would rely heavily on the contention that launch and detonation would violate the two Just-War criteria of discrimination and proportionality, although that contention seems to be fairly well supported by Kenny's argument. What is not straightforward, however, is the premise of the embedded argument, viz., that the intention to retaliate is immoral. Although this premise is the cornerstone in many popular arguments against nuclear deterrence, and may claim some intuitive appeal, it stands in need of more formal support which is not forthcoming.

Premise (4*) is a clearer statement of (4), since it brings to the fore the only acceptable interpretation of that premise. But of course without the acceptability of (1*), the entire argument is doomed.

2.3 The Deterrer's Intentions

In the relatively few pages which he devotes directly to deterrent intention, Kenny focuses his attention on three basic questions. The first of these asks if the intention to retaliate is actually present in a standard policy of nuclear deterrence.[23] Equating that intention with the threat to retaliate, Kenny claims that it is present as the means to achieving the ultimate purpose of deterrence, i.e., peace and national security. To his credit, he is careful to separate out the two intentions of deterrence, the conditional intention to launch an attack, and the 'ultimate' intention to deter attack. Few commentators on nuclear policy, especially those critical of deterrence, are able to make that distinction.

Secondly, he asks if the intention to retaliate must necessarily be a part of the threat of deterrence. His answer is 'probably not.'[24] All that is needed is the availability of the nuclear option, and 'a willingness which consists in preserving their use as a genuine option.'[25] While this may suffice for an effective deterrent, it does not, as Kenny rightly notes, affect the moral argument against deterrence. For if the wrongful intentions principle is acceptable, then surely so is a wrongful willingness principle. Such a policy would still involve, in the words of Finnis & co., 'a murderous will.'[26] And despite Kenny's claim, it is not clear that mere willingness does not itself include the intention, for there is already a 'conditional intention expressed in those threats' which the willingness empowers.[27] Thus the issue of whether the intention to retaliate is a necessary part of deterrence may have no real bearing on the morality of the policy.

---

[23] Kenny, 1985, p. 49.

[24] Ibid., p. 50.

[25] Ibid.

[26] Finnis, et al., p. 111.

[27] Ibid., p. 112.

Kenny's final question follows from the first two: If the intention to retaliate is present (but not necessarily so), what role does it play in deterring potential aggression? His answer seems to be that the intention plays no significant part. Twice quoting McGeorge Bundy, Kenny clearly implies that, of the two constituent parts of deterrence, the existential component, not the declaratory policy, does the real work.[28]

Not only is the declaratory policy superfluous and ineffective, it is the locus of the immorality of deterrence, since there can be 'no credible and rational declaratory policy to be enunciated as the justification of nuclear weapons.'[29] His point is that the immorality of deterrence lies in policies which set out the grounds and methods for launch and detonation. The implication is that, without the declaratory policy, deterrence might be morally acceptable (at least as an interim measure), a view which seems to be shared by the U.S. Catholic Bishops and Pope John Paul II.[30] It is also a view which is fundamental to Kenny's later suggestion for a timetable for disarmament, which includes both a renunciation of declaratory policies and the continued maintenance of an interim (existential) deterrent.[31]

---

[28] Ibid., pp. 51 and 52.

[29] Ibid., p. 51.

[30] See United States Catholic Conference, paras. 173-74, for a discussion of Pope John Paul's much-quoted, but enigmatic, statement on deterrence: 'In current conditions 'deterrence' based on balance, certainly not as an end in itself but as a step on the way toward a progressive disarmament, may still be judged morally acceptable.'

[31] There are critics of deterrence who disagree that the morality question can be addressed exclusively to declaratory policy. Finnis & co. (p. 127), for example, argue that the hardware of the existential deterrent carries with it an inherent threat every bit as powerful, and immoral, as any declaratory policy: 'Each side's capabilities and deployments, apart from any articulation of threats, are intended to create a threat that the other side will assuredly be *made to* suffer destruction in a nuclear engagement.'

2.4 Kenny's Contributions

The Logic of Deterrence offers several important contributions to the corpus of philosophical thinking about deterrence. First, the general tone of Kenny's arguments is something which those on both sides of the nuclear debate should welcome. He successfully follows a path of objectivity and rationality between the extremist views of nuclear buildup and unilateral disarmament. Using an acceptable Just-War foundation, and despite some nagging problems (such as equating disproportionate and intentional killing of the innocent, and unquestioningly embracing the escalation hypothesis) he often argues convincingly for that middle ground.

Secondly, Kenny offers a highly credible plan for disarmament, one which successfully balances strategy and morality. While a detailed discussion of his plan is outside the scope of this thesis, it would be negligent to pass over his work in this area without mentioning it, along with the fervent hope that such a plan might some day soon find its way onto the negotiating tables at Geneva.[32]

Kenny's third contribution of significance, and one which is more germane to the present topic, is his ability to distinguish and separate out similar but critically distinct aspects within deterrence. He takes the opportunity to demonstrate this ability on three different occasions. The first is his distinction between threat and intention. Although he eventually equates the two, he calls attention to the fact that they are not one and the same. The second distinction he makes is between the existential and the declaratory deterrent. Noting that the hardware is separate from any declared intention to use it serves to show that deterrent intention may be subject to independent examination.

The final distinction which Kenny draws is also the most relevant. He identifies

---

[32] The 1987 signing of the Intermediate Nuclear Forces Reduction Treaty by the United States and the Soviet Union, which implemented the first step of stage (3) of Kenny's Unilateralist Agenda (1985, p. 70), gives some indication that ideas such as his are being turned into practical policy.

the conditional intention to launch an attack as a separate and independent part of the ultimate intention to prevent aggression. This analysis represents a major first step toward a complete understanding of the moral difficulties of deterrence.

Although Kenny's approach to the issue is laudable, his argument that the inner intention to retaliate is wrong and should be abolished if one wishes to repair the morality of deterrence, falls short. For he fails to take the next key step in that analysis and see that the inner intention, while immoral on its own, may be justified within the larger context of deterrence. The examination of the acceptability of that claim is reserved until Part V.

3. KEMP: INAPPLICABILITY OF THE WRONGFUL INTENTIONS PRINCIPLE

Few writers have ever attempted to construct a deontological defense of nuclear deterrence. Although none has wholly succeeded, Kenneth Kemp perhaps comes the closest to putting together an acceptable argument.[33] He begins by laying out the primary premises of what he believes is the strongest deontological argument against deterrence. These include the Principle of Discrimination (that deliberate killing of the innocent--noncombatants--is immoral), the Wrongful Intentions Principle (that it is wrong to intend what it is wrong to do), and what he calls the Fact about Deterrence (that nuclear deterrence requires an intention to kill the innocent).[34]

---

[33] Kemp, 1987a, pp. 276-97.

[34] This argument bears a strong resemblance to Jeff McMahan's 'Deontologist's Argument', (esp. pp. 517) although Kemp's criticism--as well as his conclusion--differs dramatically from McMahan's. The argument, as McMahan puts it, is that deterrence 'involves a conditional intention to use nuclear weapons in ways which would be immoral. Because it requires this intention, which is itself held to be wrongful, nuclear deterrence is deemed to be immoral, even if it is successful and nuclear weapons are never used.' McMahon rejects this argument in favour of a similar, but less absolutist argument which replaces the wrongful intentions principle with an alternative principle, that 'it is wrong, other things being equal, to risk doing that which it would be wrong to do.' (p. 535). In this he is in agreement with Dworkin (p. 450) that risk analysis provides a sufficient method for judging deterrence. However, McMahan is liable to the same criticisms I raised with Dworkin. Indeed, the fact that McMahan overlooks the significance of autonomous effects is even more damaging, since his principle is only *prima facie* binding.

While Kemp finds no problem with the Principle of Discrimination, he raises doubts about the truth of the remaining premises, and therefore the acceptability of the conclusion that nuclear deterrence is immoral.

Thus his defense of deterrence is indirect. Rather than constructing a deontological argument supporting it, Kemp tries to show that no such argument has been (can be?) found against deterrence. Although he fails on at least one occasion to offer adequate support for his claims, it is nonetheless instructive to examine his arguments, especially where they incorporate his understanding of deterrent intention.

3.1 Must Deterrence Threaten Innocents?

Kemp's first attack against the argument comes in the form of denying the factual premise, that inherent in any effective policy of deterrence is an intention to kill noncombatants. Kemp argues that such an intention is neither conceptually nor factually linked to western deterrent policy. As to the first, deterrence, even simply deterrence by threat of punishment (as opposed to deterrence by threat of denial), need not require an intention directly to kill innocents. It simply requires an 'intention to inflict some kind of damage' to some sufficiently valued asset of one's enemy.[35]

This seems to be correct. A deterring agent need not threaten to inflict a level of damage which will outweigh that which he himself might suffer; he need only threaten a level of harm which will outweigh the potential gains of his opponent's aggression. As we saw in Chapter 2, the deterrent threat affects the practical reasoning of the agent's opponent, altering his analysis of the benefits and costs of aggression. The threatened retaliatory harm must merely tip the scale against aggression.

---

[35] Kemp, 1987a, p. 279.

3.2 The Special Nature of Deterrent Intention

Kemp moves on to question the applicability of the Wrongful Intentions Principle to nuclear deterrence. While he readily admits that the principle can be applied in ordinary cases, where the intention is the 'last step on the way to the performance of the action itself,'[36] he questions whether it can be applied to deterrence. He argues that it cannot, since deterrent intention is unique in ways which are significant for the applicability of the Wrongful Intentions Principle.

Deterrent intention is not like an ordinary intention (as we defined it in Chapter 3) because it is not the 'last step' on the way to action. Unlike an ordinary intending agent, the deterring agent uses the intention to isolate himself from its execution. This uniqueness is not simply due to the fact that the agent only reluctantly forms the intention. For although this is (probably) true, it is not morally significant. Nor is it because the deterrent intention is conditional, since this also has little impact on the question of the morality of the intending agent. Indeed, Kemp argues that deterrent intention should not even be classified as conditional in the standard sense (where fulfilment of the conditions is seen to lie outside the agent's sphere of control), since the deterring agent believes that fulfilment, or at least nonfulfilment, is precisely within his control. Instead, says Kemp, the uniqueness lies in the fact that deterrent intention is 'self-frustrating'. The agent believes that formation of the intention itself 'will assure that the conditions under which the immoral action was to be carried out will not arise.'[37] The intention acts as a sort of barrier which prevents, rather than facilitates, execution. It is this unique self-preventive nature of the deterrent intention which renders condemnation of the agent by appeal to the Wrongful Intentions Principle problematic at best.

---

[36] Ibid., p. 288.

[37] Ibid., pp. 289-90.

There are several difficulties with Kemp's understanding of deterrent intention. The first of these has to do with his claim that the intention is only *prima facie* wrong, and thus may still be formed in the course of fulfilling other, more stringent duties. His argument for this claim seems to rely on an appeal to the agent's previously unmentioned intention to deter aggression.[38] Up to this point, the reader has been led to believe that the intention in deterrence is to inflict some kind of damage on the enemy. But there is apparently another aspect of intention which is pivotally important in proving that the original intention (which he now calls the 'conditional intention') is at most *prima facie* wrong. Although he seems to be correct, Kemp's sketchy appeal to a different intention leaves one with the impression of *deus ex machina,* and is at any rate not well suited as a defence against those critics who claim that the intention of deterrence is absolutely wrong.[39]

A further problem arises from Kemp's claim that execution of the intended act (i.e., retaliation) is purposeless. While this claim is important for rescuing the morality of the deterring agent, it leads to a serious question about the feasibility of deterrence: How is it possible for a rational, moral agent to intend to perform an act which he has no reason to perform, and indeed many good reasons not to perform? That is, given that performing x is irrational, how can a rational agent intend x? The answer seems to be that he cannot.

Kemp is silent on how to resolve this problem, most likely because he does not see it as a problem, given his earlier disputation of the Fact about Deterrence. For if retaliation does not necessarily (or even actually) involve the intentional killing of the innocent, then it may not be any more irrational than other (justifiable) acts of war. However, this seems to be inconsistent with his claim that the act is purposeless, since retaliation would have as much purpose (within the context of war) as any other

---

[38] Ibid., p. 291.

[39] See, e.g., Finnis, et al., pp. 291-94; and Anscombe, 1968, pp. 186-205.

act of war. What he is lacking is a justification for forming the intention independent of the act. While he hints at such a justification in his discussion of the other aspect of deterrent intention (i.e., deterrence of aggression), he never clarifies what that justification might be. So the problem of irrationality remains to be answered. After laying out a revised notion of deterrent intention, I shall return to this crucial problem in Chapter 11.

Despite some gaps in his argument, Kemp has succeeded in bringing forth a number of key features of deterrent intention. These features (self-frustration, lack of purpose of the intended act, the existence of another aspect of the intention) cluster around his argument that deterrent intention is unique among intentions, and therefore exempt from evaluation under the Wrongful Intentions Principle. Taken together, they signify the necessity of an independent moral analysis of deterrent intention, and point the way to a revised understanding of that intention, one which will begin to clarify the real moral issues in deterrence.

4. KAVKA: THE PARADOXES OF DETERRENT INTENTION

In his 1978 paper, Gregory Kavka identifies three paradoxes of deterrence which arise directly out of attempts to determine the moral status of an agent's conditional intention to retaliate.[40] The paradoxes are disturbing, he says, because they result from applying certain widely accepted moral doctrines, which he calls bridge principles (so-called because they link evaluations of actions and agents), together with a foundation of utilitarianism, to a typical (although perhaps not actual) deterrence situation. Kavka's goal in pointing to these paradoxes is not to condemn deterrence as morally unacceptable, but to call into question the underlying moral principles, arguing that they ought to be revised or qualified.

---

[40] Kavka, 1978, pp. 285-302.

Kavka begins by setting out two assumptions about nuclear deterrence. The first of these represents the state of affairs, which he calls a Special Deterrence Situation (SDS), wherein nations might have reason to form deterrence policies. He stipulates the definition of an SDS by reference to four conditions: (1) The intention to retaliate is genuine; (2) The intention does in fact deter aggression; (3) Large and roughly equivalent negative utilities are associated with both aggression and retaliation; and (4) the deterring agent has conclusive moral reasons to retaliate.[41]

4.1 The Wrongful Intentions Principle

This situation (which he admits may not accurately reflect any real political situation[42]), taken together with his normative assumption of foundational utilitarianism,[43] leads to three disturbing moral paradoxes. His claim is that each of the paradoxes follows from the assumptions and the application of several readily acceptable moral *bridge principles* which serve to 'link together the moral evaluation of actions and the moral evaluation of agents (and their states) in certain simple and apparently natural ways.'[44] While all three paradoxes deserve in-depth examination, I shall concentrate here on the first, which Kavka states as:

> (P1) There are cases in which, although it would be wrong for an agent to perform a certain act in a certain situation, it would nonetheless be right for him, knowing this, to form the intention to perform that act in that situation.
>
> (P1') In an SDS, it would be wrong for the defender to apply the sanction if the wrongdoer were to commit the offence, but it is right for the defender to form the (conditional) intention to apply the sanction if

---

[41] Ibid., p. 287.

[42] Although Kavka is fairly noncommittal about the realism of his SDS, admitting that the paradoxes generated from it may only be of 'theoretical interest', it is clear that he believes he is accurately reflecting reality, especially if one accepts, as we have done, the practical necessity of forming a real intention to retaliate. See p. 287.

[43] Ibid., p. 287.

[44] Ibid., p. 286.

the wrongdoer commits the offence.[45]

It is the instantiation (P1') of the existential claim (P1) which is the heart of the deterrence problem for Kavka. How can one rightfully intend what cannot rightfully be done? To clearly show the paradox here, he makes explicit the first of his implicitly acceptable bridge principles, which he calls the *'wrongful intentions principle:* To intend to do what one knows to be wrong is itself wrong.'[46] It appears that (P1) is a denial of this principle, so one or the other must be rejected.

Here we have the opportunity to closely examine the wrongful intentions principle, which both Kemp and Kavka accept without formal argument. For his part, Kavka offers some evidence why we already accept the truth of the principle:[47]

(1) We consider an agent who intends to commit a wrongful act, but is frustrated in his attempt, as 'just as bad' as an agent who succeeds in performing a similar act.

(2) We view a man who changes his mind before committing a wrongful act as having corrected a 'moral failing or error'.

(3) It is 'convenient' to treat a prior intention as the beginning of the act itself.

Kavka considers these three statements to be self-evident truths. But perhaps they deserve closer scrutiny. As to (1), it is not clear that such a judgement about the moral equality of the two agents can made without more information. Kavka's description of the frustrated agent as one who is prevented from acting 'solely by external circumstances' is incomplete. There are various stages at which this agent could have been frustrated, and thus perhaps various moral judgements which can be made about him. Consider four cases:[48]

---

[45] Ibid., pp. 288 and 290.

[46] Ibid., p. 289.

[47] Ibid.

[48] I am indebted to Jonathan Glover for suggesting these distinctions.

(a) The agent contemplates the wrongful act, but is prevented from forming the intention (i.e., developing a feasible plan), as might be the case for an agent who is institutionalised;

(b) The agent forms the requisite intention, but is stopped before he can implement any part of the plan;

(c) The agent contemplates the act, forms the intention, and begins to implement his plan, but is stopped before completing his project;

(d) The agent carries out his intended plan, but is unsuccessful, such as a would-be murderer who shoots and misses.

It is at least an open question whether any of these agents is 'just as bad' as one who actually commits the wrongful act. But it seems fairly clear that the same moral judgement cannot be made in all cases of act frustration. It is certainly true that most legal systems, which very often have roots in associated moral systems, would not judge agent (a) to be 'just as bad' as agent (d).[49] Kavka's example of a frustrated agent, 'a man whose murder plan is interrupted by the victim's fatal heart attack,'[50] offers no help in clearing the confusion, since that example could apply to (b), (c), or (d), or even to (a), depending on how one defines a 'plan'.

It may be argued that the reason for the legal distinction (e.g., the unenforceability of punishing mere contemplation) is not derived from the system's moral roots. But while this argument may be effective in countering a claimed moral basis for distinguishing the bare intention of (b) from the attempted wrongdoing of (d), it does not account for the apparent moral difference between an agent similar to (a) who considers acting but resists the temptation even to form a plan, and an agent similar to (d) who succumbs to the evil temptation, but is simply an ineffective wrongdoer.

---

[49] Kenny (1966, pp. 648-50) discusses the theoretical underpinnings of making a legal distinction between two such agents.

[50] Kavka, 1978, p. 289.

There is a further problem with this piece of evidence. At first blush, it seems that the claim is inconsistent with Kavka's consequentialist normative assumption. To condemn the first agent with the second not only disregards the actual negative utility wrought by the second, but also ignores the possibility of reform, of which Kavka makes explicit use in statement (2). At most, Kavka may claim, as he admits in his subsequent discussion of conditional intentions, that the first agent is *nearly* as bad as the second.

It may be that this criticism of inconsistency makes an unwarranted assumption about the strictness of Kavka's foundational consequentialism, viz., that it leaves no room for judgements about agents apart from the consequences of their acts. And in fact, given his later rejection of act-utilitarianism,[51] it would seem that Kavka wants to retain the right to make such judgements.[52] But if this is so, it reveals a more serious problem about the claim of the agents' moral equality, and the relevance to the wrongful intentions principle. For it seems that statement (1) is true only if we sever our judgement of the two agents from our judgement of (the results of) their actions. That is, in order to condemn the frustrated agent along with his successful counterpart, we must ignore the diverse outcomes of their actions, and indeed the diversity of their actions (or inactions). Unfortunately, the truth of (1) is gained at the expense of its relevance to the bridge principle which attempts to show the connection *between the agent and his act.* Since (1) is true only if the agent can be conceptually severed from his act, it cannot function as a supporting premise for the claim that the wrongful intentions principle is true.

Alternatively, if Kavka wants to support that principle, he must allow for the relevance of the act, and its consequences, in the assessment of the agent. But this

---

[51] Ibid., p. 300.

[52] In this regard he follows Mill, (p. 18, n2) who argues for the retention of an ability to make a 'moral estimation of the agent', based on his motive, or 'habitual disposition.'

allowance raises serious doubts about the truth, rather than the relevance, of (1). And here again, it seems that the most Kavka can claim is the diluted judgement that the first agent is nearly as bad as the second.

As to statement (2) concerning the agent who changes his mind before acting wrongly, while it is probably true, there is a question about its admissibility in an argument supporting the wrongful intentions principle. It does seem to be that we judge such an agent to have rectified his moral character. But since this is a judgement about character, it has no direct bearing on the connection between acts and intentions. It might have an indirect relevance if we had some mechanism for linking acts (and intentions) to agents. However, this is circular reasoning, since (2) is purported to give credence to the very principle whose purpose it is to show that connection between act and agent.

It seems then that only statement (3), that it is 'convenient' to treat a prior intention as the beginning of the act itself, has direct relevance to the wrongful intentions principle. But it too is problematic. In that final statement, Kavka anticipates Finnis and others who see intentions as integral parts of acts.[53] By association (or perhaps more properly, by the fallacy of division) intentions are judged to have the same moral status as the actions of which they are a part. But we have seen in Chapter 3 that this view of intention is unacceptable. While ordinarily the relationship is so close that one may judge them as a unit, there are cases in which the intention ought to be judged apart from its associated action.

It must be noted that these problems with Kavka's evidence do not lead us immediately to a rejection of the wrongful intentions principle. Rather, they point out that more justification is needed. But at the very least they show that on the basis of the evidence presented the principle cannot support the transfer of an absolute prohibition from an act to its intention. That is, the claim that doing x is absolutely

---

[53] Finnis, et al., p. 80.

wrong cannot be used in conjunction with the principle to support the further claim that intending *x* is absolutely wrong. At best the two premises may support the weaker claim that intending *x* is *prima facie* wrong. And even this claim is doubtful without more convincing support for the principle itself. Until that support is found, the question of the acceptability of the wrongful intentions principle must remain open. It might also be worthwhile to recall that Kavka's project here is to develop an argument for the revision of all such principles because of their inapplicability to deterrence, not *vice versa*. Given that goal, it may be to his advantage to shy away from conclusive reasons for their acceptability.

4.2 Rejection of the Bridge Principles

If we assume the acceptability of the wrongful intentions principle, and accept Kavka's understanding of deterrent intention, it seems that he has discovered a genuine dilemma, the pattern of which is repeated in his other two paradoxes. In each case, a seemingly intuitive moral principle leads to bizarre and counter-intuitive results when applied to the case of nuclear deterrence. And in each case, Kavka strongly suggests that the problem lies not with deterrence, but with the underlying principle.

Thus in the final section of his paper he mounts an explicit attack on those commonly accepted bridge principles which are 'shown to be untenable by the paradoxes of deterrence.'[54] The principles which bridge the gap between agent and action do so at the expense of 'significantly deforming one or the other.'[55] Clearly for Kavka the principles stand in need of some qualification.

But rather than suggest what that qualification should be, Kavka examines and rejects two alternative solutions, act-utilitarianism and what he calls *extreme*

---

[54] Kavka, 1978, p. 301.

[55] Ibid., p. 299.

Kantianism, an agent-oriented system which attributes moral significance exclusively to features such as character and state of will (with apparently no regard for consequences). Kavka rightly states that both moral theories are too one-sided to accurately reflect our common moral beliefs, but he fails to follow that up with a suggestion for a compromise solution.

5. NOVAK: DISSECTING DETERRENT INTENTION

One possible solution to the problem of deterrent paradoxes is that the wrongful intentions principle is applicable to deterrent intention, but to a deterrent intention whose unique nature is properly dissected and understood. That some dissection is necessary was suggested in 1983 by Michael Novak, who led a group of lay Catholics in writing an essay on the issues of nuclear war and deterrence.[56] It was written primarily as a response to the U.S. Catholic Bishops, who were at that time drafting their pastoral letter on the same issues. At that early stage in the Bishops' preparation, it appeared that they would come out in favor of unilateral disarmament.[57] Novak and his contributors sought to counter that sentiment with an alternative Catholic view.

In the process of developing a consequentialist defence of nuclear deterrence, Novak argues that the term 'intention' in the context of nuclear deterrence is equivocal, and goes on to identify three distinct uses of that term: the fundamental moral intention, the secondary intention, and the architectonic intention. This dissection of what is normally thought to be an homogeneous concept constitutes

---

[56] Novak, 1983.

[57] Although the final version of the pastoral is more compromising than the earlier drafts, calling merely for 'no first use' (e.g., §153), some commentators still maintain that the Bishops' position is tantamount to unilateralism. See, e.g., Finnis, et al., pp. 389-90; and Wohlstetter, 1983a, p. 16: 'The final [pastoral] letter rules out any use of nuclear weapons, first, second, or ever.'

Novak's most significant contribution to the study of deterrence. Unfortunately, his arguments for this split view lack much needed support. But it is nevertheless instructive to examine those arguments, and to learn where they succeed as well as fail, for in the successes lie the seeds of a better understanding of deterrent intention.

Novak introduces the topic by analysing three cases of deterrent intention which can accompany carrying a firearm, each possibly analogous to nuclear deterrent intention:

> The policeman intends deterrence but no actual use unless governed by justice and the disciplines of his profession; the burglar intends only a threatening and conditioned use outside justice; the murderer intends not a conditional but a willful use.[58]

As might be expected, Novak finds that the intention in deterrence is analogous to that of the policeman, but not to either that of the burglar or that of the murderer, although he gives no reasons in support of his conclusion. One can only guess that the operative criterion is the requirement to be governed by justice and discipline. Certainly this is crucial, especially if being governed by justice means acting so as to protect peace and just order. But it seems that this is only a necessary and not a sufficient criterion for successful deterrence. A policeman also succeeds (when he does) in deterring potential criminals with his firearm because he threatens to shoot *the criminal*. This is disanalogous to many forms of nuclear deterrence (e.g., France's *anti-cités* policy), which succeed by threatening not only the guilty, but the innocent as well, a fact which seems to have been overlooked by Novak.[59] This crucial point of disanalogy (in addition to explaining why opponents of nuclear deterrence are not thereby anarchists) points to the unique nature of the deterrent intention. There are no situations analogous enough to deterrence in the relevant respects to provide the

---

[58] Ibid., p. 62.

[59] Finnis & co. argue that all forms of nuclear deterrence suffer from this moral flaw. See pp. 132-76.

moral bridge which would allow us to make a judgement about one with reference to the other.

5.1 The Species of Deterrent Intention

Novak then begin his analysis of the types of deterrent intention. He describes the overriding intention as the 'fundamental moral intention,' which is 'never to have to use the deterrent force.'[60] As proof that this is the fundamental intention in deterrence, Novak points to 'the honourable discharge of military officers, after their term of duty expires, who have succeeded in their fundamental intention.'[61] This is at best a very weak elliptical argument, especially since several other conclusions could be drawn from the stated premise, including that the fundamental intention is 'never to use the deterrent force, ever,' a conclusion contrary to Novak's position.[62]

Finnis & co. point out a further problem in expressing the fundamental intention in this way. It is not so much an intention, but a hope or desire: 'The intention underlying deterrence focuses on its purpose--survival with freedom--not on 'never to have to use the deterrent force', which merely expresses a hope for the success of the deterrent strategy.'[63] At best, we may say that this expresses a fundamental reluctance to be forced to resort to nuclear exchange, but not an intention in the standard sense, viz., one which is linked by the agent to an action.

Novak next describes the 'secondary deterrent intention,' which is 'the engagement of intellect and will on the part of the entire public that called [the

---

[60] Novak, 1983, p. 62.

[61] Ibid.

[62] An intention of 'no use, ever' is tantamount to a bluff policy, which Novak (1983, p. 63) clearly rejects (without argument) in his discussion of the secondary intention.

[63] Finnis, et al., p. 202. Barrie Paskins (p. 99) also considers and reject a similar description of the main deterrent intention.

deterrent force] into being.'[64] It is the knowledgeable and fully conscious willingness to carry out the threat implicit in the stockpiled war materiel. Novak believes that this secondary intention is absolutely essential for the effectiveness of the deterrent, for without clear demonstrable evidence of the existence of this sort of intention, an adversary might be tempted to test the sincerity of the public resolve behind the nuclear hardware. With even just the possibility of such a temptation, says Novak, the deterrent is useless, for it is 'no longer a deterrent but only an inert weapon backed up by a public lie.'[65]

In highlighting this secondary intention as a separate and essential element of deterrence, Novak may be making the distinction between the existential deterrent on the one hand, and the declaratory policy of deterrence on the other. He is arguing against many who believe that 'mere possession', viz., weaponry alone without any announced or even formulated intention to launch, is sufficient to deter potential aggressors. Thus he goes beyond attacking the feasibility of a bluff policy, which is a deliberate attempt to commit deception, to say that it is equally wrong (i.e., ineffective) to commit a deceptive error of omission by possessing a nuclear arsenal without developing the genuine intention to employ it, should the relevant conditions obtain. But despite his offhanded dismissal of a purely existential deterrent, it is not at all clear that such a deterrent might not be successful. This is especially so if the quantity and/or sophistication of the arsenal is very great, and the execution of orders to use that arsenal is routinely practised.[66] This would amount to a massive existential deterrent plus the preparation to execute, and while this comes very close to constituting secondary intention, it does not necessarily imply that such an intention

---

[64] Novak, 1983, p. 63.

[65] Ibid.

[66] Kenny (1985, esp. pp. 83-84) argues that even with a greatly reduced nuclear arsenal the existential deterrent would be sufficient to ward off aggression.

(as yet) exists.[67]

Novak lists yet a third sense of deterrent intention, distinct from either the fundamental or secondary senses already discussed. This 'architectonic' is the objective intention springing from the society which must 'generate a complex, highly rational, socially organized' system of deterrence.[68] It seems that Novak is claiming that the public, architectonic intention exists independent of the other two types. That is to say, even if all individuals involved in the deterrence policy were bluffing (viz., even if there were no secondary intention), an architectonic intention would nonetheless exist. But based on this reference, it is unclear how we are to understand this type of intention as distinct from the previous two. Finnis & co. take it to be the expression of deterrence as a public act, 'specified in and by public policy.'[69] Novak himself lends credence to this interpretation in an earlier article, where he refers to this architectonic intention as 'present even if leaders in any one administration privately and subjectively decided never to use or to threaten the use of nuclear weapons.'[70] But if this reading is accurate, then the architectonic is hard to distinguish from the secondary intention, which Novak also characterises as arising from the will of the entire public.

Another possible interpretation is that the architectonic is an extension of the existential deterrent, including not only the hardware, but the 'software' as well, viz., the plans, programs, and execution orders necessary to carry out retaliation. This extended deterrent has a corporate life of its own apart from the individuals who operate it; it is a creation of the public will, which continues to finance its operation. But while this is a plausible interpretation, it too is not far removed from Novak's

---

[67] Although approaching the problem from a different perspective, Finnis & co. (pp. 107-10) have put together a fairly sound rebuttal to the 'mere possession' argument.

[68] Novak, 1983, p. 63.

[69] Finnis, et al., p. 202. For this reason they believe that deterrence cannot even in theory be based on a bluff.

[70] Novak, 1982, p. 40.

description of the public involvement in the secondary intention. Thus it seems that this third category of deterrent intention may well be superfluous.

In reasoning towards his final position, Novak commits a number of errors. His arguments, where present, are not well laid out, and at one point his statements about the fundamental and the secondary intentions portray the two as basically inconsistent.[71] Nevertheless, his insight into the discrete nature of deterrent intention, which Kemp also hints at, is the key which may unlock the moral puzzle of deterrence. The next part of this thesis involves a further exploration of that unique nature.

[71] Novak, 1983, p. 66: 'We uphold the fundamental intention of deterrence that no nuclear weapon *ever be used*. We uphold the secondary intention of *being ready to use* the deterrent within the narrowest feasible limits, as indispensable to making deterrence work.' (emphasis added).

PART IV: A REVISED NOTION OF DETERRENT INTENTION

We try to make do with a Newtonian politics in an Einsteinian world.
--Jonathan Schell

Throughout the foregoing discussion of opposing views of nuclear deterrence, it has become increasingly clear that there is something anomalous about deterrent intention. Those who oppose deterrence have tried to gloss over this anomaly; those who support it have tried to accentuate its features. But neither side has satisfactorily grasped its nature or moral significance. It is my purpose in this part of the thesis to develop a more complete and acceptable understanding of that intention. I shall do so by highlighting the unique nature of deterrent intention, and by explaining that uniqueness by reference to a new dualistic interpretation. This interpretation will be designed to reconcile the anomalies uncovered by the critics in Chapter 5 with the standard notion of intention given in Chapter 3.

6: THE UNIQUENESS OF DETERRENT INTENTION

1. SELF-FRUSTRATION 101

2. AGENT CONTROL OVER CONDITIONS 103

3. INTENTION-PREFERENCE MISMATCH 105

4. AUTONOMOUS EFFECTS 107

6: THE UNIQUENESS OF DETERRENT INTENTION

We threaten evil in order not to do it.

--Michael Walzer

Several writers, especially those arguing in defence of deterrence, have sought to point out that deterrent intention, and therefore deterrence, defies ordinary standards of moral judgement. We have seen for example that Kemp, Kavka, and Novak all appeal to the apparent uniqueness of deterrent intention as part of their supporting arguments. To begin to clarify the nature of deterrent intention, we need to examine in more detail some of the reasons why these critics argue for the uniqueness claim. Thus, this chapter is devoted to a critique of the claim that deterrent intention is unique by examining the four primary reasons why the critics discussed in Chapter 5 make this claim.

1. SELF-FRUSTRATION

The first reason for the uniqueness claim is the realisation that deterrent intentions seem to defy the linear relationship model of intention given in Chapter 3, where an intention is 'successful', i.e., fulfills the agent's purpose, just when its

associated act is executed to produce the result which matches the agent's initial preference. This linear flow from preference through intention to result is exactly the relationship envisaged by the agent as he forms his intention to act. The linear relationship model highlights the tightly interdependent, almost indivisible, relationship between act and intention. Deterrent intentions, however, do not seem to be subject to the same criterion of 'success'. On the contrary, a deterrent intention is successful just when it *prevents* the execution of its act, or to be more precise, just when it prevents the occurrence of the antecedent conditions necessary for execution. For this reason, deterrent intentions are labeled 'self- stultifying' by Kavka and 'self-frustrating' by Kemp.[1]

It should be noted that deterrent intentions technically are not *self*-frustrating (or stultifying). A more accurate description is that the intention, or rather the publication of the intention, produces within the recipient of the deterrent threat a strong 'reluctance' to act in a way which will give rise to the conditions necessary for executing the intention. The stronger the reluctance, the more effective the deterrence policy, with an ideal deterrent producing an overriding unwillingness to act. Although the intention gives credibility to the announced threat, it is actually this unwillingness, rather than the intention itself, which prevents the occurrence of the antecedent conditions. For this reason, 'self-frustrating' might be a misleading term. But the fact remains that there is an oddity here. Whether it is a direct self-frustration or merely an extended causal connection does not seem to affect the observation that this is unique among intentions.

---

[1] Kavka, 1978, p. 290; Kemp, 1987a, p. 291.

2. AGENT CONTROL OVER CONDITIONS

The second reason why deterrent intentions appear to be unique has to do with the amount of control which the agent exercises over the fulfillment of the conditions necessary for execution. Although deterrent intentions are classified as conditional,[2] it is important not to mistake their uniqueness as resulting merely from that classification. Often a novice will begin a defence of deterrence by claiming that the intention to retaliate is not immoral because its execution is conditioned on the realisation of a certain state of affairs, and therefore the intention is not subject to evaluation by, say, the wrongful intentions principle, since the deterring agent is not directly responsible for that state of affairs. This argument of course lacks merit. We have seen (Chapter 3) that the mere existence of external conditions cannot exonerate an agent who maintains a wrongful intention. The moral character of the agent is determined by his willingness to form such an intention, not by whether that intention will be executed.

Nevertheless, deterrent intentions are unique among conditional intentions. They differ from ordinary conditional intentions in two important senses. The first has to do with the nature of the conditions. In the ordinary case, the conditions may have no more than an arbitrary relation to the result which the agent is trying to achieve. The unlikelihood of rain and my intention to go golfing are not related except in so far as I have formed the latter conditionally upon the former. However, in a deterrent conditional intention, the conditional clause refers to a state of affairs the occurrence of which the agent is specifically trying to prevent. That is, the intention

---

[2] Kemp (1987a, p. 291) argues that a nuclear deterrent intention is not conditional at all. It does not indicate 'what the agent would be willing to do, even what he would be willing to do under unusual or unexpected circumstances,' since the conditions are not meant to be fulfilled. Nevertheless, execution of a nuclear deterrent intention *is* conditioned upon the occurrence of a particular state of affairs (namely that state which the agent is trying to prevent), and thus may be properly classified as conditional.

is aimed at undermining the realisation of the conditions. Thus the relationship between the intention and the conditions is not merely arbitrary. Indeed it may be argued that successful deterrence is the direct result of the causal connection between the deterrent intention and the nonrealisation of the state of affairs which constitute those conditions.

The second sense in which the two types of conditional intentions differ regards the relative probabilities of the conditions occurring. Whether the relevant conditions obtain is ordinarily outside of the agent's sphere of control. It is in a sense a matter of luck since that occurrence is independent of the agent's intentions. To put this another way, the influence is uni-directional for ordinary conditional intentions: The agent's actions are affected by, but do not affect, the realisation or non-realisation of the antecedent conditions of his intention.

But this is not the case with deterrent conditional intentions, where the influence is, and is designed to be, bi-directional. In a deterrence situation, the agent seeks to dissuade another from acting to produce the antecedent conditions of his intention. He forms and announces his intention with the expressed purpose of exercising a sort of negative control over the relevant state of affairs, viz., his forming the intention exerts a causal influence on the *non*realisation of the conditions. While the deterring agent may be affected by the fulfillment of the conditions (or may not be--indeed at that point he would cease to be a *deterring* agent), he certainly seeks to affect that fulfillment.[3]

Deterrent conditional intentions form the basis not only of nuclear deterrence, but also of other interpersonal (and often adversarial) situations, such as many parental

---

[3]The notion of bi-directional influence bears a passing resemblance to the macro-economics Hypothesis of Rational Expectation that participants in the market place do not make systematic mistakes, which was put forth against the traditional view that market forces operate independent of the participants. For a discussion of this hypothesis, see, e.g., Begg, esp. p. 29.

sanctions issued in the course of child-raising. They also underpin the legal justification of criminal punishment (accepting that the deterrent theory of punishment explains at least part of the reasoning behind the sanctions[4]). Ideally in these cases, the real threat (i.e., the intention) to punish is not the last step before action, nor is it meant to be. Indeed, a sanction would be considered perfectly effective if it was never required to be imposed.[5]

3. INTENTION-PREFERENCE MISMATCH

There is, according to at least two critics, a third aspect in the nature of deterrent intention which distinguishes it from the ordinary variety, one which jars our common notions about preferences and intentions.[6] As we have seen, intentions primarily arise from preferences, either fundamental or instrumental. And because of the close tie between intention and preference, little importance is attached to the distinction between intending and preferring to intend.[7] But in the case of deterrent intentions, there is a significant gap between the two: the agent prefers having the deterrent intention, but also has a strong (but apparently not overriding) preference *not* to act on that intention, even though he accepts a certain risk that he will do so. That is, the agent simultaneously holds (a) a preference for the intention; (b) a strong preference to withhold the intended action; and (c) no counterbalancing preference to act on the intention.

---

[4] For a discussion of the theory, see, e.g., Airaksinen, pp. 66-74; Mabbott, pp. 39-40; and Hart, esp. pp. 133-34.

[5] This is the reasoning behind the claim that nuclear deterrence has been perfectly effective in keeping the peace in Europe for over 40 years.

[6] Kavka, 1978, p. 281; Kemp, 1987a, p. 288. In discussing this aspect, I shall not attempt to enter into the debate over the acceptability of the desire-belief model of intentionality. For discussions of its acceptability, see e.g., Davidson, esp. pp. 91-102; and Bratman, esp. pp. 6-8.

[7] Kavka (1978, p. 291) makes this comment with respect to intending and desiring to intend.

Although it points us in the right direction, emphasising the intention-preference mismatch has several problems. First, it is not clear that an agent can in fact rationally hold (a), (b) and (c) unless preferences have no bearing at all on intentions. For it seems that the three positions jointly held imply a bluff, or to use parallel construction, the preference for a bluff. The strong preference against acting, which presumably is tied to the conclusive moral reasons against acting, coupled with no preference to act, means that the rational agent would have absolutely no preference to act. In the face of that, it is difficult to believe that he would nevertheless want to have the intention.[8] Here we encounter the crucial and perplexing problem of the rational formation of intentions to which I shall return in Chapter 11.

A second problem for this grounding of the uniqueness claim is the overly heavy emphasis which it places on the moral significance of the agent's desires (as a species of his preferences). We have seen in Chapter 3 that desire cannot be made to play such a foundational role. At issue is not whether the agent *wants* to commit the wrongful act, but whether he, however reluctantly, *intends* to do so. Kavka and others who use this distinction are open to the criticism that 'desire', or lack thereof, does not impact on the question of morality in the same way as intention. That is, an agent is not exonerated from the evils of his intended deed simply because he did not want to so act. But although the emphasis is misplaced, pointing out the mismatch represents a recognition of the division within deterrent intention, a division which I shall clarify in Chapter 7.

---

[8]It might be objected that (a), (b) and (c) are not incompatible since one may rationally hold conflicting preferences. I may for example want to go for a walk now and want to sit and write a letter now. However, this objection fails to appreciate the crucial difference between desire and preference. A preference, as the predominant or overriding desire, and the result of rational reflection, is subject to consistency restraints which are not applicable to simple desires. I cannot rationally prefer both to take a walk now and write a letter now.

4. AUTONOMOUS EFFECTS

The final reason offered in support of the claim that deterrent intentions are unique is also the most powerful. Recall that in the ordinary (linear) case, the only significant outcomes of intentions are the acts (and their results) which flow from those intentions. The only noteworthy result of my intention to see *Hamlet* was my subsequent act of going, and the results and consequences which that act produced. The intention itself did not directly produce any tangible results, apart from some minimally important (at least in terms of the present discussion) 'fixing' of my future action into some coordinated schedule.[9]

But the situation is quite different in the case of deterrent intentions, which produce what Kavka calls 'autonomous effects that are independent of the intended act's actually being performed."[10] In deterrence, the effect is the instilling in another a disposition (the 'reluctance' discussed above) against acting to produce the antecedent conditions of the intention. These autonomous effects provide reasons for forming the intention which are separate and distinct from any reasons for (or against) acting on that intention. Thus they are morally significant, and indicate the need for an independent moral evaluation of the intention apart from the act itself.

The identification of the existence of autonomous effects as an integral part of deterrent intention provides the clearest evidence yet for the uniqueness claim. It seems that these effects play the central role in any genuine policy of deterrence. The deterrer does not wish to provoke his opponents into forcing his hand; on the contrary, his aim is prevention. He sees that an avenue to that goal--whether or not

---

[9]Bratman (esp. pp. 15-17) focuses on this 'fixing' function of intentions, and its stabilising role in his planning theory. While such a function is important in that context, it seems to play a negligible part in distinguishing between ordinary and deterrent intentions.

[10]Kavka, 1978, p. 291.

it is the only, or even an acceptable, avenue--is by announcing his intention to respond, since this (announcement of) intention produces the preferred effect, viz., a reticence in his opponents to test the resolve inherent in his intention.

That deterrent intentions produce autonomous effects does not of course resolve the moral problems of nuclear deterrence. But it does begin to indicate where those problems lie. In particular, the existence of these effects gives evidence to the hypothesis that deterrent intentions may not fit the standard interpretation of intention given earlier. At the very least, it suggests that they require a closer examination.

7: A DUALISTIC MODEL

1. THE COMMON CONCEPTION OF DETERRENCE 111

2. A DUALISTIC INTERPRETATION 112

   2.1 Composition and Function of the Intentions 113

3. THE ANOMALIES RESOLVED 116

   3.1 A Reexamination of Kavka's First Paradox 118

4. REMAINING PROBLEMS 119

7: A DUALISTIC MODEL

Although the writers we have been examining have succeeded in identifying certain anomalies in the concept of deterrent intention, none has been able to describe the concept in a way which can satisfactorily account for the noted peculiarities. Kavka, for example, feels compelled to reject the standard (and intuitively acceptable) bridge principles which produce his paradoxes. It may well be that he can see no other alternative because he has failed to more accurately analyse the nature of the uniqueness. Indeed, it may be that the principles need not be rejected at all, but will easily apply to a deterrent intention which is more clearly understood. I shall in this chapter set out an alternative, dualistic analysis of a general notion of deterrent intention which will be most obviously applied to nuclear deterrence, but which will also clarify the nature of intention in all deterrence endeavours. For nuclear deterrence in particular, this analysis may not only point the way to a reconciliation of deterrence and the bridge principles, but also lay the foundation for a moral defence of that policy.

1. THE COMMON CONCEPTION OF DETERRENCE

Any adequate theory of intention which seeks to explain the notion with respect to deterrence should, at the least, square with our intuitions about it. So before beginning this analysis, it might be helpful to recall the common understanding of intention in a genuine policy of deterrence. As a general schematic representation of such a policy, let us call an endeavour genuinely deterrent just when an agent A seeks to deter another agent B from performing a certain type of act *y* by threatening to respond to *y* by performing act *x*, where *x* would result in unacceptable negative consequences for B. The purpose or goal of such an endeavour is not the performance of *x*, but the deterrent prevention of *y*. A maintains his deterrent capability not to threaten to do *x per se*, but to defend against the threat of *y*. This purpose is reflected in Kemp's observation that once the project of deterrence has failed, viz., once *y* has occurred, 'executing the intention does nothing to further that project.'[1] At that point, the actual execution of *x* is purposeless, at least with regard to the project of maintaining deterrence.

Given this purpose, what is commonly seen to be the intention in a deterrence endeavour is the intention to act so as to prevent *y* by deterring B. And the act which prevents *y* is the announced formation and maintenance of a real threat (i.e., a conditional intention) to perform *x* if *y* occurs, backed of course by the necessary preparation for carrying out that threat. This gives the first indication of the true nature of deterrent intention: For the 'act' intended seems to be not a *bona fide* act, but a further intention. That is, the act intended to prevent aggression is the formation

---

[1] Kemp, 1987a, p. 292.

of a second, conditional intention to do *x*.[2] That an intention gives rise to another intention is a clear break with our linear model of intention given in Chapter 3. And the realisation of this break opens the door to a more accurate moral picture of deterrence.

2. A DUALISTIC INTERPRETATION

The 'piggyback' notion of intention spawning intention leads to the discovery that deterrent 'intention' is a misnomer; there are in fact two types of intention at work here. The initial or primary intention, which we may call I_P, is the intention to act so as to deter B from performing *y*. The secondary intention, I_S, is the intention to do *x* if conditions warrant, i.e., if *y* occurs. Each of these intentions requires a more thorough examination.

Before beginning that examination, recall where intention fits in to the linear relationship picture proposed in Chapter 3. Under normal circumstances, the preferences of an agent produce purposes, or goals, which 'ground' intentions. These intentions in turn produce intentional actions from which flow the results and consequences of those actions. To work backwards through our earlier example, the pleasure I enjoyed was a direct result of my seeing *Hamlet* at Stratford, an intentional action which flowed directly from my intention to go. That intention was produced from the earlier development of a purpose embedded in a plan, which itself flowed from my preference to go, an instrumental preference triggered by a more fundamental preference for pleasure. Here we come full circle, and may call my (complex) action successful since the result of that action (pleasure) matched the initial fundamental

---

[2] It might be argued that an intention, or its formation, is a sort of mental act, so that deterrence is not as unique as it appears. But while this might be so, it certainly requires an expansion of the common notion of 'act', in which case we may use the necessity for this expansion as evidence of the uniqueness of deterrent intention.

preference.

2.1 Composition and Function of the Two Intentions

The primary intention in our schematic representation of deterrence, I_p, is to act so as to deter B from doing *y*. This is Novak's fundamental intention, and Kemp's *deus ex machina;* it is what Kavka describes as the 'ground of the desire to form the [secondary] intention.'[3] And it is the conflation of this primary intention with the more obvious conditional intention to do *x* if *y*, I_S, which leads to the anomalies in nuclear deterrence observed by these writers (where '*y*' is 'aggressive action' and '*x*' is 'nuclear retaliation').

From whence does this primary intention arise? Beginning at the most basic level, it comes from a fundamental preference for some basic human good,[4] which then usually gives rise to an instrumental preference central to a well-defined purpose to achieve the fundamental preference by deterring B's performance of *y*. Securing this end is the goal derived from the primary intention, I_p.

Given this origin, to what does this primary intention lead? It seems fairly clear that it gives rise, not to an action, at least in the ordinary sense, but rather to a further intention, I_S, the conditional intention to do *x* if *y*. This intention is made credible by the development of execution plans or other preparations aimed at demonstrating to B that A is ready to act on I_S. And the result of this pseudo-action (i.e., the formation and maintenance of I_S) is the actual deterrence of B's performance of *y*, or more precisely, the deterrence of (some of) the very conditions necessary for the execution

---

[3] Novak, p. 62; Kemp, 1987a, p. 291; Kavka, 1978, p. 291.

[4] To look for anything more basic than this would be to engage in psychology, or at least some area of philosophy beyond the scope of this thesis.

of the act on which I_S is directed. So we may say that I_P is successful when the agent's input preference is matched by the actual result. Thus, the analysis of I_P is relatively uncomplicated.

But our analysis is only partially complete. There remains an important, perhaps essential,[5] element in deterrence, viz., the conditional intention to do *x* if *y*. This is the intention which I have labelled, following Novak, as the secondary intention, I_S.[6] At first glance, the linear analysis of I_S seems to be much more straightforward than that of I_P. The intention leads to the performance of *x*, provided the necessary antecedent conditions are satisfied. The moral difficulty for deterrence occurs when *x* is a morally wrong act, for example when '*x*' is 'nuclear retaliation', with the attendant expected consequences of nuclear war.

But as with the primary intention discussed above, we need to complete the analysis by examining the goals and preferences behind this secondary intention. Here it becomes clear that I_S is not as ordinary as it first appeared. What also becomes clear is the convoluted interrelationship between I_S and I_P, an interrelationship which has led most if not all critics (erroneously) to view them as one and the same. The main object of I_S is the performance of *x* should deterrence fail. While this seems to fit well into our developing linear model, it is the last part of this goal, '*should deterrence fail*,' which provides an important distinction. For the execution of I_S is external to the endeavour of deterrence, and therefore the act *x*, even if it is immoral, has no direct bearing on the question of the morality of deterrence.[7] Thus the immorality of doing

---

[5] Finnis & co. (pp. 104-31) offer an extensive argument in favour of the claim that actual possession of the intention to retaliate is a necessary condition for (effective) deterrence. We have seen alternative arguments in Chapter 4.

[6] Novak, 1983, p. 62.

[7] It may be argued that execution is external to the endeavour only in the case of nuclear deterrence (and then only if one accepts the escalation hypothesis that crossing the nuclear firebreak will inevitably lead to massive global exchange, annihilating--among other things--any possibility of future deterrence scenarios). In criminal law, for example, the execution of punishment is an integral part of deterrence, since it bestows credibility on the threat. But this argument lacks merit, even in the case of deterrence in criminal law. For what actually deters future crime is not past punishment, but the future intention to punish. A well-prepared and announced threat which has yet to be carried out is not therefore non-credible,

x does not lead directly to a condemnation of the deterrence policy. And although it may indirectly impact such a policy, e.g., by calling into question the moral character of the deterrer, it is not at all clear that it would serve necessarily to condemn the agent. For as this dualistic analysis makes plain, A's primary intention is to prevent y by deterring B, not to perform x. As a result, it makes more sense in judging his character to place greater emphasis on this primary intention, and not the secondary one. This of course is not to say that I_S has no moral relevance, but rather that its importance should be proportional to the role it plays in the overall process.

That the agent should be so judged is supported by the alternative, and perhaps more important role of I_S, viz., as the 'act' which flows from the primary intention, the act whose result is to ensure, as far as it is possible, that the antecedent conditions of that very intention, I_S, never occur. For it is that within this dualistic framework, I_S itself plays a dual role. First, it is a conditional intention whose associated act is x. And secondly, it functions as an 'act' whose direct result (i.e., without a further intervening action) is the deterrence of B, which is the goal of the primary intention. The secondary intention produces the preferred desired results in its role as action, not intention. Thus the dualistic model demonstrates that the uniqueness of deterrent intention lies not in the fact that it defies the linear relation model of ordinary intentions, but rather its uniqueness comes from the role of intention as act.

The interrelationship between I_S and I_P is further demonstrated by examining the preference behind the purpose which gives rise to I_S. For it seems that the instrumental preference here is not what would be expected for an intention to perform x, viz., the preference to do x (for whatever more fundamental reason). But rather the preference here is to prevent y, which is exactly the instrumental preference behind

---

but rather perfectly effective. The announced and genuine threat of, say, capital punishment for parking violations would not be non-credible simply because no one dares to park illegally (although it would fail to be what Honderich (pp. 58-60) calls an economical deterrent).

the primary intention as well. This is because the *raison d'etre* of a deterrence policy has nothing to do with actually carrying out the threatened response. Indeed, that is the very antithesis of the entire project.

As one searches further for the origins of the secondary intention, it becomes increasingly difficult to distinguish its roots from those of the primary intention. This difficulty has contributed to the failure to distinguish correctly these two distinct intentions, a failure which has resulted in the confusion regarding nuclear deterrence which we have seen in Part III.

3. THE ANOMALIES RESOLVED

Given this dualistic model, we can now clear that confusion by resolving the anomalies which puzzled the earlier writers. The first of these was the conclusion that deterrent intention is self-frustrating, a conclusion which jolted our intuitions about the nature and purpose of intentions. But given the dualistic perspective, we can see that it is not the case that deterrent intention frustrates or stultifies itself by preventing the very act intended. Rather, that prevention is the result of I_P, the primary intention to deter B's performance of *y*, not as it first appears, the result of I_S. Although the secondary intention works as a conduit in this process, and so is intricately related to that prevention, it is not directly self-frustrating.

The second anomaly we noted in Chapter 6 was the fact that the deterring agent exercises negative control over the fulfillment of the conditions of his intention, whereas normally the conditions are independent of the agent. But from this revised perspective, that anomaly also disappears. The result of forming I_P (not simply I_S) is the deterrence of B, i.e., the prevention of the conditions of I_S. So the agent does not (within one linear relationship) exercise direct control over the conditions of his

intention. The secondary intention does play a part, but only as a pseudo-act whose only purpose is to produce the goal contemplated by the formation of the primary intention.

Thirdly, the mismatch of intention and preference apparent in a monistic model is resolved from this corrected vantage point. The preferences of the deterring agent are perfectly consistent with his formation of I_P: He simply prefers to prevent *y*, which is precisely the goal of I_P. The agent considers the endeavour a success just when his input preference is matched by the outcome, viz., just when *y* is prevented. But we can also see that this preference is consistent with the formation of I_S, since that secondary intention is formed in order to achieve the goal of I_P, not in order to achieve its own intended result. The only remaining anomaly then is the fact that the preference (prevention of *y*) which gives rise to I_S is not immediate and 'fitting', but is one step removed, giving rise to I_P which in turn leads to I_S. And this is due to the alternative role of I_S as a pseudo-act.

Finally, the production of autonomous effects is readily explained by the dual role of the secondary intention within the endeavour. Ordinarily, one does not produce effects directly from intentions. Nor does this happen in deterrence: The 'autonomous' effects are simply the result of I_S *qua* act, not *qua* intention. It is the secondary intention as pseudo-act flowing from I_P which produces the results identified by Kavka as 'dominant in the moral analysis' of nuclear deterrence.[8] But, as the dualistic model shows, these effects are not autonomously produced; they are the more standard results of intention via action.

---

[8] Kavka, 1978, p. 291.

3.1 A Reexamination of Kavka's First Paradox

As a further demonstration of how the dualistic model resolves the problems of deterrent intention discussed by the writers we have examined, we may now return to Kavka's paradox which we encountered in Chapter 5:

> (P1) There are cases in which, although it would be wrong for an agent to perform a certain act in a certain situation, it would nonetheless be right for him, knowing this, to form the intention to perform that act in that situation,

which for nuclear deterrence becomes:

> (P1') In an SDS, it would be wrong for the defender to apply the sanction if the wrongdoer were to commit the offence, but it is right for the defender to form the (conditional) intention to apply the sanction if the wrongdoer commits the offence.[9]

Kavka claimed that the paradox arose from the apparent contradiction that there are (or at least seem to be) acts which are wrong to perform, yet right to intend. Under the dualistic model, however, we have come at least this far: The primary intention does not represent an instantiation of (P1). For while it is right to *form* the intention to act so as to prevent aggression, it is not at the same time wrong to *act* on that intention, since that act is merely to form the further intention, I_S.

Seen in isolation then, I_P seems to successfully pass the test generated by appeal to wrongful intentions principle. However, it may be objected that this argument for the acceptability of I_P moves too quickly, and assumes the acceptability of forming I_S, when this assumption is (as yet) unwarranted. The assumption is especially suspect if we let x be an immoral act, such as nuclear retaliation. In such cases, the secondary intention to perform x would be judged to be immoral, either by the wrongful

---

[9] Ibid., pp. 288 and 290.

intentions principle, or by something like the Davidsonian view that having an intention implies that one has made an overall value judgement that the intended act is right.[10] Such a view in cases where doing x is wrong leads to a contradiction, since the agent believes both that x is right (implied by his forming the intention to do x) and that x is wrong.

It seems indisputable that, all things being equal, the rightness or wrongness of an intention should be inextricably bound up with that of the intended act, that intentions which are judged to be wrong ought not to be formed. But all things are not always equal. There may be relevant considerations about the reasons for forming a particular wrongful intention which nullify the ordinary negative judgement about such an intention. The dualistic nature of deterrent intention provides evidence that the existence of very important results should lead to a reexamination of the dictum that it is wrong to form intentions upon which it would be wrong to act. Recognition of the need for this reexamination will at the very least cast doubt on any claims about the absoluteness of that dictum.

4. REMAINING PROBLEMS

In one sense then deterrent intention does not present an awkward problem for the model of action which I have presented. Much of the aberrant data which has worried critics can be explained and easily accommodated once the 'intention' is correctly identified and categorised into its constituent parts.

But in another important sense deterrent intention remains unique and anomalous, especially with regard to the secondary intention. Deeply embedded within the endeavour of deterrence, I_S functions in its dual role not simply as an intention,

---

[10] Davidson, pp. 7-8.

but as an act which produces preferred results. And even within its role as conditional intention I_S is unique, for it is not meant to be a last step on the way to action. Quite the opposite is true. The intention to respond by doing x does not exist in that usual close proximity to its associated action which in ordinary circumstances allows for the application of the wrongful intentions principle.

However, despite the remoteness of the intention to its action, and despite even the aversion of the deterring agent to execution, I_S remains a deep and troubling problem for the morality of nuclear deterrence in particular: It is still the murderous intention to retaliate whose execution is immoral and very likely irrational. Knowing this, how can a rational and moral agent form such an intention? The answer to this crucial question occupies much of the remainder of this thesis.

PART V: THE PRINCIPLE OF DOUBLE INTENTION

In Part IV of this thesis I have argued that, to be properly understood, deterrent intention must be dissected into its constituent parts, and that this dissection reveals the distinction between the primary intention to prevent aggression and the secondary intention to retaliate. While this analysis clarifies the nature of deterrence, it leaves us with a nagging problem: The secondary intention, despite its minor role in deterrence, is nevertheless immoral.

It would be natural at this point, and indeed proper in most situations, to condemn an agent who, knowing of this immorality, nevertheless proceeds to form the secondary intention. As we saw in Chapter 5, this was the reasoning behind the wrongful intentions principle, the purpose of which is to tie the morality of an agent to that of his action, using intention as the mortar which binds the two together. But before we condemn the agent, there is at least one possible avenue of justification to explore which may offer a reason to separate judgements about intentions from those about the agents who form those intentions.

In this section I present that avenue of justification under what I shall call the Principle of Double Intention. I shall argue that this principle, analogous (as its name

suggests) to the Principle of Double Effect, can serve as an aid for assessing the moral worth of an agent who forms two seemingly inconsistent intentions, where one of those intentions is judged to be immoral.

I shall first spend some time analysing the Principle of Double Effect. This analysis will lay an important foundation for the argument supporting the plausibility of the Principle of Double Intention. I shall argue that, while the Principle of Double Effect is admittedly open to criticism for its ability, or rather inability, to resolve borderline cases, it nevertheless contains a core of acceptable thinking which can lead to fundamentally sound moral assessments, especially when it is used to judge the goodness of agents. Drawing on that analysis, I shall argue directly for the plausibility of the Principle of Double Intention, and answer relevant objections to it. I shall then use that principle to assess the morality of an agent who engages in nuclear deterrence. Is he justified in forming the (admittedly immoral) intention to retaliate, or should we condemn him for it?

Finally, I shall close this part of the thesis by addressing the question which we have delayed from Part IV: Is it rationally possible to form the secondary intention to retaliate, given the admitted irrationality of acting on that intention? I have deferred this question until the end because its positive answer springs from the morality issues of deterrence and the Principle of Double Intention.

8: THE PRINCIPLE OF DOUBLE EFFECT

1. THE CONDITIONS OF THE PRINCIPLE 126

1.1 Act Neutrality 126
1.2 Intention vs. Foresight 128
1.2.1 'There Is No Difference'
1.2.2 'There Is No Significance'

1.3 The Means-End Relation 134
1.4 Proportionality 135

2. A METHOD FOR APPLYING THE PRINCIPLE: THE QUALIFICATION TESTS 136

2.1 The Countermeasures Test 137
2.2 The Nonfulfilment Test 138
2.3 Application of the Tests: Some Examples 139
2.3.1 Lethal Pain Relief
2.3.2 Strategic vs. Terror Bombing

3. RELEVANCE OF THE PRINCIPLE 143

8: THE PRINCIPLE OF DOUBLE EFFECT

The Principle of Double Effect (hereinafter abbreviated PDE) is a doctrine arising out of the struggle to come to moral terms with agents whose actions produce both good and evil consequences. Not specifically tied to any type of normative theory, it is a skeletal assessment tool compatible with any moral theory which 'allows that there are kinds of acts which are good and bad.'[1] Far from being non-controversial, the acceptability of the PDE is questioned by a wide variety of critics. Consequentialists lead the attack against its theoretical underpinnings, while some deontologists question its practical application.[2]

Although the origins of the PDE may be traced back to St Thomas Aquinas's discussion of killing in self-defence,[3] the first formalised statement comes from Joannes Gury in 1874:

---

[1] Boyle, 1980, p. 537. Many commentators on the PDE, especially those critical of its plausibility, argue that it is essential to (and indeed is only comprehensible within) normative systems which include absolute prohibitions. See, e.g., Anscombe, 1970, pp. 50-51; Duff, p. 68; Richards, p. 381; and Kemp, 1987b, p. 94. Against that view, Boyle argues that there is no such necessary connection.

[2] For examples of the former, see Hart, pp. 126-27; Glover, p. 88; and Sidgwick, p. 202. For an example of the latter, see Anscombe, 1970, p. 51.

[3] Aquinas, II-II, 64, 7: 'Now moral acts take their species according to what is intended, and not according to what is beside that intention.' See also Kenny's discussion of Aquinas and double effect (1973, pp. 140-41).

> It is licit to posit a cause which is either good or indifferent from which there follows a twofold effect, one good, the other evil, if a proportionately grave reason is present, and if the end of the agent is honourable--that is, if he does not intend the evil effect.[4]

Gury's point is that the evil which results from a particular act does not necessarily condemn that act, or, significantly, the agent who carries out the act. Implicit in this statement is the claim that intention plays a role in determining questions of right and good, that intention has what Joseph Boyle calls an 'act-defining character.'[5] The kind of act which produces the requisite double effect will be 'specified by the [intended] good effect as a morally good act.'[6]

The PDE is an attempt to define those instances--and common sense tells us that there are many–when an agent is justified in causing an evil effect which he would not otherwise be permitted to bring about.[7] As Boyle correctly points out, the purpose of the PDE is not merely to excuse the agent for producing the evil because of some mitigating circumstance, but rather to justify him and his action--to eliminate his culpability.[8]

Underlying the principle is the idea that the same act (leaving aside for the moment the problem of act descriptions and the distinction between acts and consequences) can be performed, not both rightly and wrongly, but by a good agent and a bad agent. Norvin Richard's distinction between an ordinary dentist and a

---

[4]Gury, p. 5, as quoted in Boyle, 1980, p. 528.

[5]Boyle, 1980, p. 531. The view of intention as act-defining has not always met with universal acceptance. See, e.g., Kenny's discussion of intention (1973, pp. 129-46), especially on Aquinas (pp. 138-40).

[6]Ibid.

[7]Throughout my discussion of the Principle of Double Effect (and its offspring, the Principle of Double Intention), I shall use the term 'act' to mean act of commission, and leave open the question of omissions and their effect on the principles, although it must be said that at least one critic of the PDE, Philippa Foot (p. 25), believes that its strength turns on its (erroneous) claim to distinguish what is done (commission) from what is allowed to happen (omission).

[8]Boyle, 1980, p. 529.

torturer using dental instruments illustrates this point.[9] Both drill into tooth enamel, an action which produces pain, but the dentist alone is justified in his action. In this case, as in others, the PDE serves to highlight a moral difference between the two agents, not necessarily between their two (identical) acts.

1. THE CONDITIONS OF THE PRINCIPLE

Following Gury, all formulations of the PDE include a set of conditions which must be satisfied before an agent can be exonerated by appeal to the principle. Although the lists and descriptions tend to vary,[10] four standard conditions can be extracted:

(1) The act itself, considered apart from its consequences, must be good or at least morally neutral.

(2) The agent must intend only the good effect; the evil, although foreseen, must not be intended.

(3) The evil effect must not be a means to producing the good.

(4) The good effect must be proportionately great enough to justify permitting the evil to occur.

1.1 Act Neutrality

The first condition of the PDE, that the act which produces the two effects must itself be good or at least neutral, is a prohibition against 'positing a cause' which is

---

[9] Richards, pp. 384-85.

[10] For instance, Gury's first condition, that the 'end of the agent is honourable,' (Boyle, p. 528), is mentioned by no other commentator except Kemp (1987b, p. 92). Aquinas does not specify any conditions, except perhaps a reference to the unintended nature of one of the outcomes ('Nothing hinders one act from having two effects, only one of which is intended, while the other is beside the intention.' II-II, 64, 7), which lends some support to J. Ghoos' claim that St Thomas did not advocate the principle as it is now understood. See Kemp, 1987b, p. 91. For other versions of the conditions, see Fisher, pp. 30-31; Frey, pp. 259-61; Ford, p. 26; and Wasserstrom, p. 153.

intrinsically wrong. This presupposes acceptance of the assumption that acts can be the proper subject of independent moral evaluation, an assumption which consequentialists may well find objectionable.[11] It also presupposes that the principle can only be effectively employed from within a normative theory rich enough to contain a set of criteria for identifying types of acts as intrinsically right or wrong, which again presents a conceptual difficulty for the consequentialist.

Jonathan Glover raises an objection at this point. How does one effectively draw the line between an act and its results?[12] Before determining if an act is intrinsically wrong, we must first be able to describe it correctly. This involves determining the boundary between the act and its results. How does one, for example, describe Sirhan Sirhan's assassination of Senator Robert Kennedy? Possibilities range from a neurological account of Sirhan causing a minor muscle contraction in his right index finger, to something much more inclusive of consequences, more of the tabloid headline form: 'Sirhan Slays Kennedy in Cold Blood.' Where along this spectrum does the appropriate description lie?

This seems to present a serious problem for the PDE. It looks as if the principle could be invoked to justify any act, provided only that the agent is clever enough to invent an acceptable act description. For example, as Glover says about the classic problem case of abortion used to save the mother, 'Killing the foetus while it is attached to the womb will be permitted under the description 'saving the mother's life.'[13]

But the PDE can withstand such an attack. For while this first condition seems

---

[11] See e.g., Frey, p. 260: 'Because, then, the [PDE] only makes sense against a backcloth of acts which are intrinsically right or wrong, it follows that no consequentialist can embrace [it].' This does not, however, imply that the PDE can only be acceptable to absolutist moral theories. See Boyle, 1980, p. 537.

[12] Glover, p. 90. Bennett (1966, p. 86) makes a similar point, although his argument stresses more the lack of moral significance in the action/consequence distinction.

[13] Glover, pp. 90-91.

open to abuse by clever agents, the principle, taken as a whole, is sophisticated enough to discriminate between authentic cases where *all* of the conditions obtain, and mere perversions of the principle. An assassin such as Sirhan, seeking to justify his act of murder by describing it as the morally neutral act of flexing his index finger, would either fail to pass the proportionality condition (4), or the intention and means conditions (2) and (3), depending upon his description of the good effect which resulted from his act. If the good was, say, isometric exercise, then he would have failed to effect a good proportionately great enough to justify permitting the evil. Alternatively, if the good to be achieved was perhaps the elimination of a young liberal force in American politics, he would have violated condition (3) by using the evil as a means to that good (assuming, contrary to reality, that such an elimination could in any way be seen as a good), as well as condition (2), since the evil would have been intended.

So while Glover's objection would be damaging to the acceptability of condition (1) taken in isolation, the remaining conditions work in concert to prevent any illicit manipulation of the principle.[14]

1.2 Intention vs. Foresight

The second condition, that the evil effect, although foreseen, cannot be intended by the agent, is at once the cornerstone and the millstone of the principle. Without this condition, the PDE would have neither force nor purpose. But it raises fundamental problems and as a result has borne the brunt of the criticism levelled against the principle.

---

[14] Kemp (1987b, p. 94) has in fact argued that this first condition, far from being damagingly problematic, may even be superfluous, at least given Gury's original conditions.

Before discussing some of the most potent objections to this condition, it is important to clarify exactly what is meant by 'intended' and 'foreseen', and to examine the difference between the two. What does it mean for a result to be intended? First, recall from the earlier discussion of ordinary intention (Chapter 3) that results are not intended *per se*. To say, in the present case, that 'agent A intends the good result *g*' is a shorthand version of the fuller, and more accurate statement, 'A intends to perform *n* in order to achieve *g*,' where *n* is a morally neutral act as specified by condition (1). This is the complete intention statement, and emphasises the linear relationship which exists between intention, act and result, which in turn allows us to speak (albeit elliptically) of an 'intended result'. The intention itself arises from a preference, either fundamental or instrumental. In ordinary circumstances, as we have seen, the entire process is considered 'successful' when the ultimate goal 'matches' the input preference.

Therefore, to speak of an 'intended goal', as distinguished from one which is merely foreseen, highlights the special linear relationship contemplated by the agent, which includes preference, intention, act and (matched) result. The agent, in linking the preference to the intended goal, adjusts his intention to 'track' that goal.[15] This sort of outcome tracking does not occur for merely foreseen results, viz., those consequences of an agent's action which he considers to be incidental or side effects,[16] permitted to occur as by-products of action.

In the case of acts of double effect, the agent may be well aware of the high probability that evil will result from his act. But that evil is not part of, and does not

---

[15] This is the idea behind Foot's comment (p. 25) on the distinction between intended and foreseen effects that the former, and not the latter, are 'aimed at'.

[16] Some proponents of the PDE (e.g., Bratman, p. 140; Kemp, 1987b, p. 92) prefer to use the term 'side effects', rather than foreseen consequences, as they believe that this renders more perspicuous the relevant distinction with intended effects. Anscombe (1982, p. 13) goes so far as to suggest that the principle be called the Principle of Side Effects. While use of this term might improve clarity, I shall continue to use the more traditional terminology.

figure into, the special linear relationship which the agent has formulated to match his preferences to his expected goal. Indeed, as I shall argue in Section 3 of this chapter, the evil result actually plays a negative role in the (moral) agent's reasoning process in that he may alter his plans in order to avoid or reduce the evil.

1.2.1 'There is No Difference'

There are many objections to the claim, inherent in condition (2), that a morally significant distinction can be drawn between intended and foreseen results. Most of these attacks can be grouped into two broad categories, those which deny any difference between intention and foresight, and those which allow that there may indeed be a difference but argue that the distinction is not morally significant.

One of the strongest statements of the first type of objection comes from Henry Sidgwick, who implies that any apparent difference between intention and foresight is chimerical.[17] When an agent chooses a particular course of action, he becomes responsible for all foreseen consequences of that action, regardless of which results he directly intends.

There are several problems with Sidgwick's attack. First, he conflates 'intention' and 'choice', incorrectly implying that the two are morally equivalent. But they are not. Although it might be, at least in some simple cases, that when an agent chooses a particular course of action he develops the intention to produce all foreseen outcomes, this apparent identity of intention and choice dissolves in more complex scenarios, or under more careful scrutiny. Choosing a scenario does not commit an agent to intend every component therein. As Bratman puts it, 'a rational agent will

[17] Sidgwick, 1962, p. 202.

normally only intend certain elements of that scenario.'[18] It is true that if I choose to pick up the pen on the desk before me, I also develop the intention to achieve the result of an elevated pen. But it is not as clear that I also intend the expense of energy or the minor deterioration of the pen casing, both of which are necessary components of my chosen course of action. Sidgwick's implied assumption that choice and intention are interchangeable stands in need of support which is not readily available.

Secondly, Sidgwick accuses PDE proponents of attempting to 'evade responsibility for any foreseen bad consequences.'[19] But responsibility is not the issue. Advocates of Double Effect are not generally trying to deny that the agent is responsible for causing evil.[20] What they do claim is that the PDE can be used to determine whether that production of evil condemns the agent; they argue that it may be the case that the agent, although responsible, is not as morally culpable for unintended effects.[21]

Other critics who deny the difference do not follow Sidgwick in subsuming under intention all foreseen consequences of action, both certain and probable. Rather they deny any difference between intended and foreseen effects when the effect is inevitable, or 'invariably and inseparably' linked to its act in such a way as to make the connection seem 'conceptual rather than contingent.'[22] An agent who denies that he intends such an effect stretches credibility beyond its acceptable limits, and

---

[18] Bratman, p. 161. Boyle (1980, pp. 535-37) offers a similar argument by reference to the relative 'voluntariness' of intended versus foreseen consequences.

[19] Sidgwick, p. 202. This sentiment is echoed by Frey (p. 265): 'He who knowingly brings about a consequence does not escape responsibility for it merely because he did not directly intend it.'

[20] Anscombe is a notable exception to this general rule. She claims, without argument, that an agent 'is not responsible for the bad consequences of good actions.' (1968, p. 200). Hers is not the standard position on this issue, nor does it seem to be supportable under any ordinary understanding of 'responsibility'.

[21] See, e.g., Kenny (1966, p. 649): 'It may well be correct to hold the agent responsible for these [foreseen] consequences, but that only means that we can be held responsible for more than we intend.'

[22] Hart, p. 123. See also Frey, p. 263.

opens himself to the kinds of attacks which Anscombe heaps upon abusers of the PDE who manipulate it into a 'perverse doctrine.'[23] To use Robert Hoffman's admittedly trivial example (at least in the moral, rather than the culinary realm), if I intend to carve a roast, and believe that I cannot do so without dulling the carving knife, then I actually intend to carve the roast *and* to dull the knife.[24]

But in usual moral parlance, 'intention' does not have such wide meaning. Those results which are said to be intended are just those results which, as mentioned above, the agent aims at, which he 'tracks' by adjusting his intentions and actions to achieve his preferred result. Surely, in this sense, I cannot ordinarily be said to intend a dull knife edge each Sunday as I set to work on the roast. Indeed, such an 'intention' would run contrary to what I do (intentionally) just before I begin to carve, i.e., *sharpen the knife.*

This response is similar to the response of Boyle and Sullivan, who use the much less trivial counter-example of an adult stutterer who foresees the inevitable result of his speech, but 'struggles against the unwanted but practically inescapable concomitants' of that act.[25] Here it is clear that the agent, far from intending the foreseen consequence, actually takes steps to prevent it. Yet without a distinction between intended and foreseen outcomes, we are forced to conclude that the stuttering is nevertheless an integral part of the agent's intentions. This counter-example also provides a convincing rebuttal to Sidgwick's argument that all foreseen consequences are intended.

---

[23] Anscombe, 1970, p. 51. See also Kenny's discussion of Pascal (1973, pp. 140-41) and the absurdity of (mere) direction of intention.

[24] Hoffman, p. 390.

[25] Boyle, 1976, p. 358.

1.2.2 'There is No Significance to the Difference'

The second major type of objection raised to the intention-foresight distinction is one which admits the difference, but argues that it is of no moral significance, amounting to at most a 'merely verbal difference.'[26] The complaint here is that the question of culpability cannot hinge on the narrow difference between knowing the results of one's action and intending those results. At a certain point, that knowledge is sufficient to condemn the agent.

Responses to this objection are not difficult to find. However, most appeal to examples which purport to show that the intention-foresight distinction does in fact have moral significance. Unfortunately it is often difficult to generalise from the specifics of each case, and the objectors have at their disposal a number of equally plausible counter-examples. The two best rebuttals are not casuistic, but argue for a sort of graduated scale of moral culpability. In the first, Boyle argues that such culpability is based on an ordered set of meanings of 'voluntary', the paradigm of which is the 'execution of deliberate, free choice.'[27] The closer human behaviour comes to this paradigm, the more it is regarded as the subject of moral evaluation. The intended results of one's action mirror this paradigm in a way that the foreseen consequences do not. This difference, based as it is on the notion of 'voluntary', is morally significant.

Kemp follows a similar line of argument, but bases his scale on the notion of responsibility, and concludes that 'one is more responsible ... for what one intends

---

[26] Boyle, 1980, p. 533. For other statements of this problem, see Duff, pp. 73-74; Richards, pp. 385-87; Hart, p. 127; and Glover, p. 88.

[27] Ibid., p. 534.

than for what one merely permits.'[28] This notion of relative responsibility (where 'responsibility' is understood to mean 'culpability', and not merely 'accountability') is a doubly effective response, since it serves to blunt not only the criticism against the significance of the distinction, but also the earlier attack by Frey and Sidgwick (§1.2.1 above) that proponents of the PDE are trying to evade agent responsibility for foreseen effects. On the contrary, supporters of the principle argue, with Kemp, that the agent is simply *less* responsible (i.e., culpable) for the foreseen than for the intended results of his action.

Much of the force of the objection of no moral significance results from an erroneous view of the preemptory role of desire (as a species of preference) in distinguishing intention and foresight. But while desire does play a role in intention formation, what is central to the difference between intention and foresight is the agent's fundamental disposition toward the effects. Desire is only a part of this disposition; how the agent 'tracks' each of the effects is another part of his disposition. What is needed then is a method to determine the true nature of that disposition. In section 3, I shall offer a series of tests designed to help make that determination.

1.3 The Means-End Relation

The third condition of the PDE, that the evil effect must not be a means to producing the good, stirs nearly as much debate as the second. The point of the condition is to emphasise the evil as a side effect of the act, as a by-product which serves no purpose for the agent in achieving his intended result. The condition is often explained in terms of the immediacy of the good effect, that the good must flow directly from the act. Unfortunately, this explanation tends to place overly heavy emphasis

---

[28] Kemp, 1987b, p. 69.

on the temporal or causal sequence, neither of which, claim the opponents, has moral significance.[29]

However, concentrating an attack on the significance of forbidding evil means is both inappropriate and ineffective. Such an attack misses the whole point of the PDE, which is to determine the moral status of the agent whose act produces the two effects. Thus the relevant factor is not the causal sequence *alone* (otherwise there would be only one qualifying condition), but rather the way in which the agent carries out his act, the disposition of the agent toward his effects. The causal sequence is relevant only as one aspect of that disposition.

It may be that an overly inflated emphasis on the causal sequence has led not only to unfair criticism of the PDE, but also to misapplication of the principle by some of its supporters. This may for example account for the problematic distinction between killing and letting die in the borderline abortion cases.[30]

1.4 Proportionality

We come to the final condition of the PDE, that the good effect must be proportionately great enough to justify permitting the evil to occur. Gury describes this condition as the requirement that 'there must be a grave reason for positing the cause.'[31] What constitutes a grave reason is unspecified, but it seems that satisfaction of this condition is determined by weighing the relative (negative) values of not

---

[29] See, e.g., Frey, pp. 280-83; Hart, p. 124.

[30] Long a staple of PDE opponents, these cases attempt to draw a distinction between directly killing the foetus (usually as the only way of removing it from the womb) and simply removing the foetus and letting it die. For a sample of the criticism of making such a distinction, see, e.g., Bennett, p. 83; Duff, p. 68; Frey, p. 268; Foot, pp. 20-21; Glover, pp. 89-90; Hoffman, pp. 391-93; and Richards, p. 394. It could well be that some supporters of the PDE have erred. Given the relevant facts, perhaps such cases should be judged together. The application tests discussed in §3 below should go a long way in helping to determine the correct use of the PDE here.

[31] Gury, p. 8, as quoted in Boyle, 1980, p. 528.

producing the intended good and permitting the evil.

This condition is straightforwardly consequentialist, calling for an evaluation of the expected outcomes of one's action. The criteria for evaluation are not stipulated, although they are presumably utilitarian rather than, say, egoistic. But the procedure for determining acceptably proportionate effects, like the procedure for defining good and evil for the neutrality condition, is a matter for the encompassing theory, not a matter for the PDE itself. Thus we shall pass over assessing the acceptability of this condition.

2. A METHOD FOR APPLYING THE PRINCIPLE: THE QUALIFICATION TESTS

Most of the objections to the PDE have tended to result from doubts about the usefulness or significance of the principle in helping to assess moral culpability, and ultimately, agent worth. They have criticised the ability to make the very fine distinctions between an agent who merely foresees the evil and an agent who in fact intends that evil, or between an evil which is an unavoidable by-product of action and an evil which is a necessary means to some end. Although I have suggested some answers to these objections, I believe that the majority (or perhaps all) of the important criticisms raised can be blunted by clarifying how and when an agent may be exonerated under the principle for bringing forth evil, and this in turn can be done by clarifying the agent's fundamental attitude towards the effects he produces. A procedure for formulating the criteria for justification can be embodied in two distinct qualification tests, the Countermeasures Test and the Nonfulfilment Test. I shall discuss each

test in turn, and then apply them to some of the more troublesome borderline cases of the PDE.[32]

2.1 The Countermeasures Test

Many of the difficulties of the PDE cluster around the relationship of the agent to the evil effect. Was it really part of his plan, intended as a means or for some other purpose? The Countermeasures Test is a first step toward a satisfactory answer about that relationship. The test asks

> Does the agent adopt all reasonable means to mitigate the evil, or to reduce the probability of its occurrence?

This in effect tests the agent's sincerity in his attitude toward the evil. Mere regret is too closely aligned to the kinds of abuses of the PDE attacked by Anscombe.[33] The Countermeasures Test goes beyond regret to determine if the agent sought to avoid the evil to the maximum extent possible (even if he considered it inevitable) by taking concrete steps against it.

The test goes to the heart of the intention-foresight distinction. For if in fact there were no difference between the two, the agent's adoption of countermeasures against the (intended) evil effect would be contradictory, since he would be intending both to bring about the evil and to prevent its occurrence. It makes explicit the force

---

[32] I am grateful to Kenneth Kemp for the early development and discussion of these tests. He has recently written on them (1987b, pp. 67-72), although he limits their use to determining the distinction between intended and foreseen results. I have tried to enlarge their applicability to the whole question of agent justification, both for the PDE and, in the next chapter, the Principle of Double Intention. Both Kemp and Fried (pp. 24-27) have suggested a third test, the Counterfactual Test, to further determine the agent's attitude. While there may be some advantage to also posing this hypothetical version of the Nonfulfilment Test, the differences between the two are so minor that the additional test may serve to confuse rather than clarify.

[33] Anscombe, 1970, p. 51.

behind Boyle and Sullivan's counter-example of the adult stutterer.[34] That the stutterer *struggles against* what he knows to be inevitable is an indication that he is taking all available countermeasures to prevent the evil effect, an indication which offers convincing evidence of the inherent difference between intention and foresight.

It is important to note that the test does not necessarily require the adoption of all possible countermeasures; reasonability alone is required. As a moral assessment tool, it can be used as a sort of grading scale. An agent is exonerated of moral culpability for producing evil just to the extent that he employs reasonable methods of mitigating that evil. One who makes a perfunctory attempt at countermeasures cannot then appeal to this test for vindication.[35]

2.2 The Nonfulfilment Test

The second test should lay to rest any lingering doubts about the agent's attitude toward the evil he has produced. The Nonfulfilment Test can be formulated:

> What would the agent do if, contrary to expectations, his action did not produce the evil effect?

This test is designed to determine if, in Hart's words, the evil effect 'constituted at least part of [the agent's] reason for doing what he did.'[36] The assumption underlying the test is that if the evil were actually intended, either for its own sake or as a means to the good, the agent would try again to achieve that effect. This same assumption

---

[34] Boyle, 1976, p. 357.

[35] Note that the test does not resolve the rather legalistic problem associated with 'moral negligence,' viz., the extent to which an agent should be aware of available countermeasures, and other questions analogous to the determination of legal negligence. These of course must be resolved before the test can become completely viable, but they are of only tangential importance to this thesis.

[36] Hart, p. 121.

also underlies the earlier notion of 'tracking' a result, viz., adjusting one's actions to achieve the intended outcome. If the evil were intended, the agent would track that result, repeatedly attempting to achieve it despite failure. Conversely, if the evil were merely an unwanted but (presumably) unavoidable side effect, the agent would not only *not* try again to achieve it, but probably rejoice in its nonfulfilment. Upon completing his drilling work on a patient who had experienced no pain, a dentist genuinely concerned with avoiding pain for his patient certainly would not continue to drill simply for the purpose of uncovering an open nerve. Not realising the bad effect will not lead the dentist to conclude that he has failed. He will instead consider the operation a great success, since he has matched his preference with the result of his action, and avoided the evil side effect in the process.

This test represents an improvement over the conditions as originally stated in that it bypasses the issue of the role of agent desire or preference, a point of contention for many critics. As the test points out, the relevant issue is not whether the agent preferred the evil effect, but whether the evil was an integral part of his plan. It plumbs the depths of the agent's commitment to the good effect, while at the same time determining his attitude towards the foreseen evil. This determination is crucial to our final moral assessment.

2.3 Application of the Tests: Some Examples

Although the two tests are not difficult to understand, it is in applying them to the tough cases of Double Effect that one can truly appreciate their effectiveness in determining an agent's culpability. Therefore, I shall apply them to two pairs of cases,

the first presented by Hart concerning injections for pain relief, and the second by Bratman about effective bombing techniques.[37]

2.3.1 Lethal Pain Relief

> A distinction must be drawn between the case where a drug is given and the patient ceases to feel pain, but as a further consequence his death is accelerated, and the case where he ceases to feel pain because a drug has been administered to kill him as the only way of saving further pain.[38]

The distinction between these two cases is at the centre of the predominant criticisms of the PDE, i.e., the intention-foresight distinction and the means-end relation question. Demonstrating an effective method for applying the principle here should do much to blunt that criticism. Let us first clarify the differences.

> Case (1): Doctor A injects drug *x* into his patient in order to relieve pain, but knows that *x* will accelerate his patient's deterioration, and thus bring on his death more quickly.
>
> Case (2): Doctor B injects drug *y* into his patient in order to kill him and thereby relieve his pain.

Is there a moral difference between these two doctors? Hart argues that there is not, since 'the overriding aim in [both] of them is the same good result, namely . . . to save human suffering.'[39] Proponents of PDE argue that there is a morally significant difference. The aim of Doctor A is pain relief; the aim of Doctor B is the death of the patient (which will then result in relief of pain).[40]

---

[37] Although this exercise is only tangentially related to the main themes of this thesis, it is nevertheless worthwhile. It will also serve to strengthen my argument for the plausibility of these tests, and consequently for the acceptability of the Principle of Double Intention presented in Chapter 9.

[38] Hart, p. 122.

[39] Ibid., p. 124.

[40] This of course assumes that pain is in fact relieved upon death, an issue well beyond the scope of this thesis.

The qualification tests should shed light on this controversy. Instantiation of the Countermeasures Test in (1) yields this question: Does Doctor A do everything he can to reduce the probability of his patient's hastened death? Presumably, yes. While details are sketchy, there is no reason to suppose that A refrains from employing all available measures to lessen the side effects. Application to (2) yields: Does Doctor B do everything he can to reduce the probability of his patient's (speedy) death? The answer of course is no. Since B is tracking his patient's death, it would be contradictory for him to also act to prevent that death.

The Nonfulfilment Test offers similar results. (1): If, contrary to the expectations of Doctor A, his patient did not die quickly, would he then try again (in some other way) to achieve that death? Of course the answer is no. On the contrary, he would probably be overjoyed and thankful to discover that the drug's expected side effects were not experienced by his patient. This is in marked contrast to (2): If, contrary to the expectations of Doctor B, his patient did not die quickly (thereby relieving his pain), would he then try again (in some other way) to achieve that death? Certainly he would, as he sees death as the only means to end the suffering of his patient. If he determined that the injection proved to be ineffective in producing the sought-after result, B would consider his action to have failed, and would try again.

These tests then highlight the difference in attitude between the doctors toward the deaths of their respective patients. It is a difference which all but the strictest consequentialist will admit has (at least some) moral significance, for it gives strong evidence of the moral character of each doctor.

2.3.2 Strategic vs. Terror Bombing

> Both Terror Bomber and Strategic Bomber have the goal of promoting the war effort against Enemy. Each intends to pursue this goal by weakening Enemy, and each intends to do that by dropping bombs. Terror Bomber's plan is to bomb the school in Enemy's territory, thereby killing children of Enemy and terrorising

> Enemy's population. Strategic Bomber's plan is different. He plans to bomb Enemy's munitions plant, thereby undermining Enemy's war effort. Strategic Bomber also knows, however, that next to the munitions plant is a school, and that when he bombs the plant he will also destroy the school, killing the children inside. Strategic Bomber has not ignored this fact. Indeed, he has worried a lot about it. Still, he has concluded that this cost, though significant, is outweighed by the contribution that would be made to the war effort by the destruction of the munitions plant.[41]

This example from Bratman is more germane to the general topic of this thesis than the preceding example, and it includes (at least in the case of Strategic Bomber) a situation in which all four of the PDE conditions seem to be satisfied.

Is there some significant difference between the two bombers? Many consequentialists deny a difference, pointing to the death of the children as evidence. The actions of the two ought to be judged together, as indicated by their known results. They are both responsible for their actions, and that alone is sufficient to determine culpability. A proponent of the PDE would disagree. Based on relative intention, Terror Bomber, far more than Strategic Bomber, is subject to moral reprobation.

Once again, the two tests may be helpful in ferreting out any relevant distinction between Strategic Bomber, Case (3), and Terror Bomber, Case (4). The Countermeasures Test gives us the following question for (3): Does Strategic Bomber adopt all reasonable countermeasures to reduce the probability of the children dying? Presumably he does. The fact that he has worried about their deaths implies that he will have done all he can (e.g., ensuring bombing accuracy, choosing attack times which do not correspond with school sessions, etc.) to reduce the probability and number of innocent deaths. In Case (4): Does Terror Bomber adopt available countermeasures to reduce the probability of the children dying? Certainly not. Reductions in probability and numbers of innocent deaths would be counterproductive for Terror Bomber. Those deaths are his means for achieving his preferred final outcome. As

---

[41] Bratman, pp. 139-40.

with Doctor B above, it would be irrational for him (assuming, without justification, that rationality is one of Terror Bomber's attributes) to adopt measures which reduce the likelihood of achieving the means to his preferred end.

The Nonfulfilment Test can also be applied to the bombers with predictable results. In Case (3): Would Strategic Bomber plan and fly another mission if, contrary to his expectations, the children escaped injury? No. He would have no reason to do so, if he had already achieved the means to his preferred end (i.e., destruction of the munitions plant). Rather, he too would be thankful for the sparing of innocent lives. In Case (4): Would Terror Bomber plan and fly another mission if, contrary to his expectations, the children escaped injury? Absolutely, unless in the interim he had changed his mind about the efficacy or advisability of his plan. In both the literal and figurative sense, he is tracking the deaths of the children as a means of terrorising Enemy. If despite his best efforts the children have escaped, he will have '*failed* to do something he was trying to do.'[42]

In this pair of cases, as with the doctors above, the tests have confirmed the claim of PDE supporters that Terror Bomber, to a much greater extent than Strategic Bomber, should be condemned for his intended action, despite their parallel results. Again, this is not to say that Strategic Bomber is not responsible for the deaths he has caused. Rather, it is to say that there exists an important moral factor which separates Strategic from Terror Bomber, one which might be ignored without appeal to the PDE.

3. RELEVANCE OF THE PRINCIPLE

The qualification tests provide a plausible solution to the problems of the PDE. They offer a way to discriminate between intended and merely foreseen consequences

---

[42] Ibid., p. 148.

by providing an 'acid test' of the agent's true, overriding intentions in a manner which renders that difference morally significant. They highlight the important distinction between evil as a means and evil as a side effect. And they prevent the worries about possible abuse of the principle, discussed by Anscombe, in which an agent 'withholds' his intention to escape culpability.[43]

Despite this buttressing, the PDE remains a controversial principle. Certainly strict consequentialists will continue to question both its theoretical foundation and its efficacy. Also problematic is its application to those borderline cases which critics are wont to offer in rebuttal. But in spite of these problems, there remains a fundamentally solid core of sound moral thinking within the principle, which among other things lends plausible support to the claim that there is sometimes a difference between intended results and foreseen consequences.

It is that solid core which I shall use in the next chapter to develop the analogous Principle of Double Intention in a way which avoids the pitfalls of the PDE. If this exercise proves to be successful, we shall be well on the way toward a more complete understanding of the moral significance of the dual roles within deterrent intention.

---

[43] Anscombe, 1970, p. 51.

9: THE PRINCIPLE OF DOUBLE INTENTION

1. RESTATEMENT OF THE PROBLEM OF DETERRENT INTENTION 146

2. THE PRINCIPLE OF DOUBLE INTENTION 148

2.1 The Scope and Purpose of the Principle 148
2.2 Initial Objections 151

3. THE CONDITIONS OF THE PRINCIPLE 153

3.1 Endeavour Neutrality 153
3.2 Primacy of the Good 154
3.3 Nonfulfillment of the Evil Intention 156
3.4 Proportionality 160

4. THE QUALIFICATION TESTS 162

4.1 The Countermeasures Test 163
4.2 The Nonrequisite Test 164
4.3 Application of the Tests: The Vulnerable Terrorist 166

5. IMPORTANCE OF THE PRINCIPLE 168

9: THE PRINCIPLE OF DOUBLE INTENTION

And much I grieved to think how power and will
In opposition rule our mortal day--
And why God made irreconcilable
Good and the means of good.

--Percy Bysshe Shelley

As there are difficult moral problems regarding acts with both good and evil effects, there are also equally difficult problems regarding agents who must undertake endeavours which require developing and maintaining both good and evil intentions. Although situations involving what we might call double intention are not as prevalent as those involving double effect, they provide some of the most troubling moral problems of our time, and therefore demand our careful scrutiny.

1. RESTATEMENT OF THE PROBLEM OF DETERRENT INTENTION

The analysis in Part IV revealed two distinct types of intention at work in a policy of deterrence, I_P, the primary intention to prevent aggression, and I_S, the secondary intention to retaliate if conditions warrant. I_P did not present any moral difficulties;

I_S, although subsidiary to I_P, did play a significant and troublesome role in deterrence. The problem of this intention can be summarised:

(1) I_S is a necessary and integral part of the intention to deter.

(2) I_S cannot be formed or maintained by a rational, moral agent.

(3) Therefore, the intention to deter cannot be formed or maintained by a rational, moral agent.

(4) The intention to deter is a necessary requirement of any successful policy of nuclear deterrence.

(5) Therefore, a rational, moral agent cannot engage in a successful policy of nuclear deterrence.

Refutation of the argument seems to lead us into a dilemma which springs from a question about the genuineness of I_S. For if I_S is not genuine, then deterrence is based on a bluff, and we must rely on the acceptability of the atomistic bluff theory. Alternatively, if we wish to maintain that the secondary intention is real, as many critics of deterrence aver and as we assumed for the sake of argument in Chapter 2, then it seems that we are forced to accept, because of premise (2), that the deterring agent is immoral. This conclusion arises from the fact that his intention to do wrong will be executed (by our definition in Chapter 3) unless that action is impeded. The only course open seems to be to deny the truth of (2), a formidable task.

Alternatively, one could argue that while (2) may be true if considered by itself, it may be false when examined within the larger context of an agent's overall endeavour. That is, while I_S may be immoral, it is not necessarily the case that the agent who forms such an intention is to be condemned. In Chapter 8 I argued that there are cases in which an agent can, without condemnation, act to produce consequences or side effects which ordinarily would have been wrong to produce. In this chapter I shall examine the possibility that it may also be morally acceptable for an agent to form and maintain an intention for which he would otherwise be condemned.

That is, there may in some cases be an affirmative answer to Gerald Dworkin's question, 'Can it be moral to commit oneself to actions which, independent of the policy in which they are embedded, are immoral?' assuming that 'to commit oneself to action' means 'to intend to act.'[1] The vehicle for assessing an agent's justification for forming such an evil intention is the Principle of Double Intention.

2. THE PRINCIPLE OF DOUBLE INTENTION

As the name implies, the Principle of Double Intention (hereafter referred to as the PDI) has its origins in the Principle of Double Effect. It is a moral principle which may be used to evaluate agents who both intend good and intend evil within the same endeavour. The PDI can be stated as follows:

> In any endeavour which requires both good and (intrinsically) evil intentions, an agent is justified in forming and maintaining the evil intention provided that the overall goal, as defined by the good intention, is morally acceptable and undertaken for a grave reason, and that acting on the evil intention is not part of the endeavour.

2.1 The Scope and Purpose of the Principle

Admittedly, the number of complex human actions to which the PDI would apply is small. To begin with, it applies only to endeavours. As I am using the term, 'endeavour' stands for 'a complex series of actions, requiring the formulation of multiple intentions, designed to achieve a singular overall goal.' Subsidiary actions within the endeavour will be done according to Anscombe's second sense of 'intention', i.e., 'with a further intention' of achieving the overall goal.[2] Endeavours are similar

---

[1] Dworkin, p. 457.

[2] Anscombe, 1966, §1.

to Bratman's plans, viz., complex structures of goal achievement requiring deliberation and both intra- and interpersonal coordination,[3] and to what Kemp calls human enterprises, i.e., 'composite[s] of human action.'[4] Examples of such endeavours are not difficult to find. Indeed, most higher level human 'actions' are of this type. To borrow from our earlier example, my evening at Stratford involved intentions and actions ranging from the purchase of tickets and choice of attire to the selection of a parking spot and the route home. The endeavour, although 'defined' by the primary goal (and thus the primary intention) of going to the theatre, nevertheless included all of these subsidiary intentions and actions.

Some endeavours will encompass apparently fundamentally opposed intentions of good and evil. Within this group will be not only deterrent intentions, discussed below, but also those acts of double effect with which many critics of the PDE took issue, viz., those in which it was not clear that foresight and intention could be justifiably delineated, especially where the foreseen effect was inevitable.[5] In those cases it may be possible for a supporter of the PDE to admit the lack of distinction and argue instead for the acceptability of the agent's action by appeal to some form of the PDI.

Additionally, the PDI will apply only to certain types of intentions, i.e., deterrent conditional intentions. As we laid out in Chapter 3, conditional intentions differ from their ordinary counterparts only in that the intending agent (passively) awaits the fulfillment of a set of conditions before he acts on his intention. We are therefore justified (discounting for the moment questions about the remoteness of the conditions) in judging agents who form conditional intentions in the same way as we judge agents

---

[3] Bratman, pp. 2-3. Bratman's theory of intention centres on this notion of planning.

[4] Kemp, 1988, p. 126.

[5] See, e.g., Kenny (1966, p. 651) where he notes that 'the rationale of the law's interest in intention lapses' when the consequence (rather than just the result) is certain to follow.

who form non-conditional ones. The existence of the conditions has no moral bearing.

But, as we discussed in Chapter 6, within the category of conditional intentions, there is a subclass of *deterrent* conditional intentions, in which the conditions, or to be more precise, the agent's attitude towards the fulfillment of those conditions, may well affect our moral judgement of him. Deterrent endeavours are those in which the agent forms an intention to achieve a certain result *r* by conditionally intending (and announcing that intention) to do *y*, where *r* (usually) represents maintaining the *status quo,* and where *y* represents some sort of punishment to be visited upon the potential disrupter of the *status quo.*[6] In short, a deterring agent is one who attempts to influence another's behaviour by threat. Accepting the deterrent theory of punishment in law as at least part of the reason for legal sanctions (viz., that the purpose of punishing individual criminals is to ensure as far as possible that the law is kept in the future), one can see that the power of the threat of criminal punishment is at its most effective when there is no need to carry out that threat. For it is the publicised *threat* of punishment, and not the punishment itself, which deters.[7]

Finally, within the class of deterrent endeavours are a small group the execution of whose threatened sanction is immoral. Most prominent among this last type is the standard policy of nuclear deterrence. As we have seen, carrying out the secondary intention to retaliate would be wrong. The existence of such an intention in an ordinary endeavour is sufficient to condemn it. However, in some situations, there may be a justification for a wrongful intention. Determining just those situations is the function of the Principle of Double Intention.

The purpose of the PDI is to aid in evaluating agents, not merely their acts, or

---

[6] Airaksinen, p. 67.

[7] See, e.g., Mabbott, p. 40: 'It is publicity and not punishment which deters.' Hart (p. 78n) makes a similar distinction 'between the efficacy of (1) the *threat* of punishment and (2) the *actual* punishment.' For a discussion of the wider issue of deterrence in punishment, see Honderich, esp. pp. 51-65.

even their endeavours. We will have occasion to examine endeavours in detail, but not simply for their own sake. Rather we shall be interested in endeavours as products of agents; our focus will be on determining the overall culpability of those who have carried out the endeavours.

2.2 Initial Objections

At this early stage of the examination of the PDI, it may be wise to discuss two objections which can already be raised against it. First, one might deny that there are any situations in which an agent is justified in forming an evil intention. It seems that there are two reasons why such a view would be held. The first of these is absolutist with respect to intentions, viz., that forming an evil intention is always wrong, regardless of any extenuating circumstances. It would seem that such an absolutist view would not extend from something like the wrongful intentions principle, which judges an intention by reference to its associated act, since as we have shown in Chapter 5 (§4.2), that principle is not potent enough to transfer an absolute prohibition from act to intention. Rather, it would result from the sort of view which holds intention to be a species of action, and absolutely prohibits all immoral acts. However, this view is subject to the same types of criticism which damage all forms of moral absolutism, e.g., that the rigidity of resulting moral rules is often incoherent or leads to contradictory judgements.

The other reason why one might claim that an agent is never justified in forming an evil intention stems from the opposite view that intentions are not wrong (or right) in themselves, but are so because they are associated with an action which is judged to be wrong (or right). This view would accept some form of the wrongful intentions principle, and argue that since one should never act to produce evil, neither should

one intend to so act.

Two responses can be made to this version of the argument against forming evil intentions. First, in order to reach the strong claim in the objection, viz., that it is *always* wrong to form evil intentions, this argument must have, like the one above, an absolutist foundation. As such, it is subject to the same sort of criticism. Secondly, one may question whether the wrongful intentions principle is acceptable, especially as it applies here. For if an act is wrong because it produces evil, it might be argued that forming an evil intention is not wrong (in the same manner), since it does not (directly) produce evil. It cannot be further objected that the only purpose in forming an intention is to produce its associated act. As we have seen in the case of the secondary intention in nuclear deterrence (Chapter 7), there may very well be reasons for forming an intention (e.g., its autonomous effects) which are independent of the act.

The second objection to the PDI which can be raised at this early stage attacks the feasibility of the principle. It accepts the possibility that evil intentions may be justified, but denies that the criteria for such justification can in fact be fulfilled.[8] Before assessing this objection we need to deepen our understanding of the principle and its conditions.

3. THE CONDITIONS OF THE PRINCIPLE

The justification conditions of the PDI can be extracted from the statement of the principle given earlier:

(1) The overall objective of the endeavour must be morally acceptable.

(2) The good intention must be primary; the evil intention, even though it is necessary for the endeavour to succeed, must be secondary.

---

[8] This objection is similar to what may be called 'just war pacificism,' which accepts the traditional Just-War Theory, but denies that the justification conditions could ever be met. See, e.g., Hauerwas, pp. 100-102.

(3) Execution of the evil act which is the object of the secondary intention must not be a requirement for achieving the objective of the endeavour.

(4) There must exist a grave reason for undertaking the endeavour.

Satisfaction of each of these conditions is necessary for an endeavour to be justified under the principle. But, as with the Principle of Double Effect, the conditions are not jointly sufficient to compel an agent to undertake an endeavour. Once again, we are dealing in the realm of permissibility, not obligation.

3.1 Endeavour Neutrality

The first condition of the PDI requires that the endeavour in question be morally acceptable, 'endeavour' being defined earlier as a 'a complex series of actions, requiring the formulation of multiple intentions, designed to achieve a singular overall goal.' To be morally acceptable, that overall goal, i.e., the purpose of the endeavour, must be good or at least morally neutral.

Defining goodness and neutrality is not a proper function of the PDI, thus an objection of incompleteness similar to that raised against the PDE at this point will fail to damage the principle in any important way. The PDI is a skeletal aid for agent assessment, and is therefore (by definition) incomplete without reference to some supporting moral theory. Objections as to how moral goodness and neutrality should be defined must be addressed to that theory, not to the principle alone.

The neutrality condition represents an improvement over that of its parent principle in two ways. First, unlike the traditionally accepted first condition of the PDE, this first condition of the PDI reflects Gury's statement (of Double Effect) that 'the end of the agent is honourable,'[9] since an endeavour, even more so than

[9] Gury, p. 5, as quoted in Boyle, 1980, p. 528.

an act, cannot accurately be viewed apart from the agent who plans and executes it. Furthermore, since assessment of an act independent of its consequences is not required, this condition is not open to the attack that such an assessment cannot be made, a problem which Glover and Bennett raise about the PDE.[10] What should be assessed is the overall objective of the endeavour, something which is much more easily discernible.

3.2 Primacy of the Good

The second condition of Double Intention, that the good intention must be of primary importance, provides the most fertile ground for generating objections to the principle. Before exploring, and answering, some of these objections, it is necessary to clarify two key notions of the condition. The first of these involves defining what is meant for an intention to be 'primary'. As I am using the term, an intention is considered to be primary in an endeavour if it provides the *raison d'etre* for that endeavour. Using the model of intention and action given in Chapter 3, the primary intention is that intention whose contemplated result can be identified as the overall goal of the endeavour. It is the *final* 'further intention' with which the acts are carried out. In Anscombe's terminology, the primary intention 'swallows up all the preceding intentions *with* which earlier members of the series [i.e., actions] were done.'[11] This relationship is usually evident in the fact that the endeavour is 'named' by reference to that primary intention. For example, my plan (i.e., endeavour) is to 'see *Hamlet* at Stratford tonight;' my intention to do so provides a name for that endeavour.

The primary intention can be distinguished from secondary intentions, which

---

[10] Glover, p. 90; Bennett, p. 86.

[11] Anscombe, 1966, §26.

although vital to the successful outcome of the endeavour (i.e., achievement of the result of the primary intention), do not bear the unique relationship to the endeavour which the primary intention enjoys. Instead, they are embedded within the larger endeavour, and therefore take on the moral qualities of the whole. In the example above, secondary intentions include my intention to purchase tickets, to make arrangements for a baby sitter, to drive to the theatre, etc. While all of these are necessary to the endeavour, none can be considered primary in the same sense as my intention to see *Hamlet.*

The second notion which must be clarified is that of a 'good' (or 'evil') intention. Two important questions surround this notion: First, can intentions actually be labeled as good or evil, or are they instead morally neutral? Secondly, if that distinction can be made at all, how should it be done? As to the first of these questions, it seems that we must consider such a distinction possible. Since intentions are indicative of character, they can be rightly assessed as good or evil based on their positive or negative impact on the agent who develops them. Furthermore, even if this does not settle the question, I should like, for the purposes of my overall evaluation of nuclear deterrence, to grant the assumption that the distinction can be made. For if it were not so, the entire issue of whether or not nuclear deterrence is morally acceptable would for the most part melt away. If judgements could not properly be made about the morality of intentions, the rightness or wrongness of developing and maintaining the secondary intention to retaliate could not be questioned; objectors to the deterrent force would be confined to consequentialist arguments about the actual harm done by maintaining such a capability.

Accepting that the distinction can be made, the question of how it should be done, like the question of endeavour neutrality, can only be answered by reference to the larger moral theory which encompasses the PDI; it cannot be properly addressed from

within the principle. However, we may note that any theory which accepts some version of the wrongful intentions principle will also include a method for assessing intentions (if only by reference to acts). Even those theories which do not embrace that principle will presumably be able to make such a discrimination, either directly or by further reference to the results of forming a particular intention.

Assuming the acceptability of these two notions, one might object that this second condition is redundant; the first condition, which requires that the overall goal of the endeavour be morally acceptable, suffices to cover the requirement for primacy of the good intention. This is especially true since the primary intention has been defined as that with which the overall goal is identified. But although it may be simpler, and perhaps more elegant, to reduce the number of conditions to a bare minimum, doing so in this case would obscure an important aspect of the PDI: While the focus of the first condition is on the goodness of the overall objective, and therefore reflects the primacy of the good intention, the emphasis in the second condition is on the secondary nature, the *non*-primacy, of the evil intention. The two conditions taken together stress the relative value to the agent of the two intentions.

3.3 Nonfulfillment of the Evil Intention

The third condition, that carrying out the evil act intended secondarily cannot be part of the endeavour, is a crucial discriminator between moral and immoral enterprises assessed under the PDI. It serves to ensure that acceptable endeavours remain properly distanced from the contemplated evil act, even though forming and maintaining the intention to do so is required to reach the overall objective. Only the intention, and not the act intended, can be considered a legitimate part of any morally acceptable endeavour.

The prohibition against execution is grounded in the earlier discussion of Double Effect. For if execution were permitted, the endeavour would be prohibited under the PDE, since the evil act would have also been intended, not merely foreseen. Execution of evil cannot be an acceptable means to achieving a good objective.

It may be objected here that the intention to do evil cannot be judged apart from the doing of that evil, since in general, intention cannot be severed from its act. Both act and intention must be judged together, since the intention is part of that act. A form of this objection was raised earlier in the discussion of ordinary intention, brought forth by those, such as Finnis & co., who see intention as 'the beginning of the act itself,' rather than a separate entity preceding action.[12] They see a conceptual link between act and intention which binds the two so closely together that they must be assessed as one. In contrast to this view, the approach in Chapter 3 suggests that intention is a separate and distinguishable precursor to action, a component in the process which can be identified and judged independent of that process. This separation allows for the formation (but not the execution) of intentions whose associated act might be absolutely prohibited. As we have seen, such a prohibition would provide (at most) *prima facie* evidence against forming that intention, reasons which could be overridden by other, more weighty considerations, such as contemplated in the fourth condition.

A more formidable objection can be raised here regarding the feasibility of intending an act which will not (or cannot) be executed. If the evil act in question cannot be carried out, then the evil 'intention' is merely a bluff, and as such carries no force in the endeavour. While this objection cannot be accurately leveled against the third condition (which states merely that execution of the evil cannot be a part of the plan, not that it is impossible), it once again raises the rationality question, which

---

[12] Finnis, et al., p. 80.

will be the focus of Chapter 11, of whether one can possibly intend that which one cannot do. But aside from that issue, the objection does not call into question the permissibility of intending an evil act. Labeling an intention a bluff does not thereby deny the moral suitability of forming that intention. As mentioned above, there may be other compelling reasons for an agent to form an intention, despite a prohibition against execution.

At the heart of the nonfulfillment condition is the claim that the endeavour can include an intention without necessarily including the act intended. This brings into focus the objection raised specifically against nuclear deterrence which we recalled at the beginning of this chapter: If the evil intention is genuine, then the agent *means* to do wrong unless he is prevented from acting on that intention, or is somehow drawn away from it. But it seems as if, by setting up the requirement that execution of the evil intention cannot be a part of the endeavour even though the intention is genuine, we have established an impossible condition. Therefore, execution must be part of the endeavour.

This objection can take one of two forms. Either one may attack the PDI for failing to admit of the direct connection between the evil act and its intention, or else one may attack the agent who forms such an endeavour for corrupting himself by setting out to do evil. The first attack springs from the distinction we noted in Chapter 3 between side effects and consequences on the one hand, and results and goals on the other. Since the evil act is genuinely intended, it cannot be considered merely a side effect or consequence of the endeavour, but must be counted among the directly intended results, if not the goals, undertaken by the agent. As such, it is an integral part of the endeavour, and thus the nonfulfillment condition can never be satisfied.

A first answer to this objection would point out that, in situations where the PDI applies, fulfillment of the conditions leading to execution of the evil intention would

indicate that the endeavour had *failed*, placing the agent in a radically altered set of circumstances from that in which he first formed the intention. As a result, that intention may well be rethought, and possibly abandoned, or else the agent may fail to act through *akrasia*.

But this answer falls short of an adequate defence. For the changes in circumstance are merely extrinsic. The agent himself has not changed, only the facts of the situation, which are external and out of his control. Just as our moral assessment of an agent does not turn on the fulfillment of the conditions of his ordinary conditional intention, so should our assessment of the doubly intending agent not turn on the fulfillment of the conditions of the secondary intention. He has already committed himself to the deed, knowing that those conditions might be fulfilled. That alone, and not the remoteness of fulfillment, should be sufficient to judge him.[13]

A more effective answer to this objection will hark back to the uniqueness of deterrent intentions. Unlike one who forms an ordinary conditional intention, the deterring agent does not lack control over the fulfillment of the conditions of the evil intention. Rather, he exerts a sort of negative control, that is, he acts to influence those he seeks to deter in order to prevent fulfillment. So the difference between an agent who successfully endeavours to deter and one who fails in that endeavour (viz., faces the occurrence of the conditions of the evil intention) is not merely a matter circumstantial distinctions external to the two agents, but is rather a reflection of their relative competence. The first is simply better at deterrence; his act of exerting negative influence is more effective.

The second form of this general objection that execution cannot be divorced from the endeavour condemns the agent for corrupting himself by forming the real intention to do evil. An agent who conditionally intends to do x in C is committed

---

[13] Finnis & co. (pp. 104-105) make this point against this type of defence of nuclear deterrence.

to perform x if he finds himself in C. He has set himself up to act, barring the intervention of some impediment to action. But if the intention is genuine, then the agent cannot *plan* on an impediment. He cannot, for example, genuinely intend to act and at the same time know that, if the time comes, he will suffer from *akrasia*. Therefore, he cannot separate the formation of the intention from performance of the act.

The answer to this version of the objection can again be found in the uniqueness of deterrent intentions. These intentions are a class of conditional intentions which seek to influence the behaviour of others by genuine threat. But unlike ordinary conditional intentions, deterrent intentions contain conditions which are not outside the control of the agent. Indeed, the very purpose in forming the secondary intention is to attempt to prevent the conditions of execution of that intention from arising. He forms the intention not as a last step on the way to execution, but specifically to *avoid* execution. Thus in the case of deterrent intentions, one can separate the act from the intention. Indeed, that separation is the essence of deterrence. While this answer will not be adequate against a charge that the secondary intention is absolutely immoral, it is sufficient to repudiate the objection that any agent who forms *prima facie* immoral intentions is wrong.

3.4 Proportionality

The requirement that there be a grave reason for undertaking the endeavour stems from a need to ensure some sort of proportionate balance in favour of endeavouring. But the difficult question underlying this condition is, what should be weighed in that balance?

Taking the cue from Kavka's qualified normative assumption for his Special Deterrence Situation,[14] it seems that we should be looking for a favourable balance of negative utilities.[15] Given this guideline, on one side of the scale should be placed the harm resulting from not undertaking the endeavour. On the other side, two possible candidates for comparison emerge, (1) the harm resulting from *acting* on the evil intention, and (2) the harm resulting from *forming* the evil intention. Since carrying out the evil act is not a part of the endeavour (as we have seen in situations where the PDI applies, acting on the evil intention actually signifies the failure of the endeavour), it seems as if (2) would be the correct choice for comparison. Since formation and maintenance of the secondary intention is required, it is reasonable to assess the impact of that formation when enquiring after the justification of the endeavour.

It might be objected that weighing the evil intention (and not the act) is unfairly tipping the balance in favour of the endeavour, since relatively little negative utility is associated with (merely) forming an evil intention. A more realistic counterbalance to the negative utility of not endeavouring is rather (1), the harm of actually carrying out the evil intended. But while the use of (1) might tend to avoid distorted measurements, it unfairly introduces into the calculation an element which is not properly part of the endeavour, i.e., the negative utility associated with performing the evil act. The third condition of the PDI makes it clear that such performance is (indeed must be) external to the endeavour.

It is perhaps the case that one should balance the harm of not endeavouring against some aggregation of the actual harm of forming the secondary intention, plus

---

[14] Kavka, 1978, p. 287.

[15] Although negative utilitarianism may not provide the only basis on which to make the proportionality assessment (one may want to balance, e.g., respective accountability of rights preservation), it does seem to make the most sense, especially since we are considering this condition with an eye toward the application to nuclear deterrence, a policy whose failure may result in negative utilities of enormous proportions.

the harm of *risking* acting on that intention, since such risk is arguably intrinsic to forming the intention, that is, intending the evil act increases the risk that the act might actually be performed. But while this seems to provide a more accurate assessment, it is faulty for two reasons. In addition to introducing the negative utility of (in this case, the probability of) acting on the evil intention, such risk cannot be quantified in any way which would meaningfully lend itself to comparison with the harm of not undertaking the endeavour. With regard to nuclear deterrence, Finnis & co. have examined and rejected consequentialist arguments against the deterrent based on the negative utility of increasing the risk of nuclear war,[16] primarily because the value (or rather disvalue) of *risk* in that realm cannot be accurately quantified for comparison. The same type of problem arises with introducing risk assessment into the utility calculations required by the fourth condition of the PDI. But despite these problems, it seems that some allowance for the risk of execution must be factored into the balance of proportionality.

As with the first condition, the actual mechanism for determining the relative values of the compared factors is a function of the supporting moral theory, not of the PDI itself.

4. THE QUALIFICATION TESTS

In general, the objections which can be raised against the PDI are similar to those raised against the PDE. One might complain, for example, that the PDI permits an unwarranted prominence of one intention over another which is irrelevant to the moral status of an intention, and therefore irrelevant to the morality of the agent who forms that intention.

---

[16] For their examination, see Finnis, et al., pp. 207-37; for their rejection see pp. 238-72.

To counter such objections, we need a more precise method of certifying whether or not an agent qualifies for justification under the PDI. Such a method can be found in the qualification tests which mirror those for the PDE discussed in Chapter 8. These tests are designed to determine the agent's fundamental attitude toward his evil intention, which in turn should prove morally significant in assessing his moral worth. If we can show that an agent who passes these tests has justified his intentions, we shall have driven a wedge between intending evil *simpliciter*, and doing so for a higher purpose.

4.1 The Countermeasures Test

The Countermeasures Test seeks to determine what the agent actually does with respect to the potential damage of intending evil. This test asks

> Does the agent adopt all reasonable means to ensure that execution of the evil intention does not occur?

Here we are not interested, as was the case within the PDE, in a mitigation of the effect of the evil intention. In keeping with the requirement of the third condition, we are trying to determine if the agent has taken steps to reduce the probability of the intended act occurring to as close to nil as is reasonably possible.

A negative answer to the question posed by this test is strongly indicative of the true character of the agent with respect to the endeavour in question. At the very least, we may say of an agent who fails this test that he is courting further evil by placing himself in the 'near occasion of sin.' Alternatively, an affirmative answer to the question shows that the agent is attempting to contravene the ordinary preference-intention-act-result linear relationship. Although the intention is required in the endeavour, and has been formed in response to his preference (which in this case would

be instrumental rather than fundamental), he has taken steps to ensure that the evil goes no further, that the associated evil act (and thus the evil result of that act) does not come to pass.

It is important to note that the Countermeasures Test does not require that all possible steps be taken to guarantee against execution of the evil act. Those steps would include the abandonment of the evil intention, since (obviously) avoiding the formation of the intention to act renders it that much less likely that the act will occur. Rather, it simply requires that the agent use reason, within the parameters of the endeavour, to determine which methods to employ in seeking to prevent the evil act.[17]

This test actually goes beyond the requirement of the third condition, which prohibits the evil act from being a part of the endeavour. It examines the motivation of the agent to ensure that the evil act is not simply not required, but is not sought after in any way by the agent. The preference must give rise to the intention alone.

An intention is (usually) judged evil by reference to its associated act; an agent is tainted by forming an evil intention. The idea behind the Countermeasures Test is this: The culpability of an agent who forms such an intention is mitigated by any steps taken to ensure that the intention is not fulfilled. The test is an attempt to assess the extent of that mitigation.

4.2 The Nonrequisite Test

The second test operates in concert with the first to firmly establish the agent's attitude toward his evil intention:

---

[17] Allowance for reasonability is not without precedent within this field of study. Most writers on Just-War Theory take the criterion that war be considered only as a last resort to mean that all reasonable means to resolve the conflict be employed before one declares war. Without reasonability in the criterion, war could never be justified, since surrender is always a possible, although not always an acceptable, option. See, e.g., Childress, p. 435; Kemp, 1987b, p. 119; and Walzer, p. 84. Fisher (pp. 20-21) implies but does not explicitly state such a restriction.

> Would the agent still develop and maintain the evil intention if, contrary to his beliefs and expectations, it were not required for the success of the endeavour?

Like the Nonfulfilment Test for the PDE, the Nonrequisite Test focuses on the agent's commitment to the evil intention. It attempts to determine if the agent harbours reasons in favour of forming the evil intention which are independent of the endeavour. An affirmative answer to the question would indicate the existence of such reasons, and thus impugn the morality of the agent. A negative answer, on the other hand, would verify the subordinate nature of the evil intention, its maintenance being merely instrumental and subsidiary to achieving the overall goal of the endeavour.

Unlike the Nonfulfilment Test, this test does not merely examine the agent's attitude toward the evil outcome which would result from acting on his secondary intention; it does not ask, 'Would the agent form the evil intention if it were not to be fulfilled?' Such a question would be too easy to affirm, given that fulfillment of the secondary intention is not part of the endeavour. The test offers a more discriminating evaluation of the agent's character. We may use it, for example, to determine if the agent would bluff instead of forming the secondary intention, if it were determined that such a bluff would be effective in achieving the objective of the enterprise. An affirmative answer to the test question would serve to indicate the sort of independent commitment to the evil intention which would lead to a condemnation of the agent. A negative answer would indicate a willingness on the agent's part to explore ways to abandon the evil without jeopardising the endeavour, a willingness very much to his moral credit.

4.3 Application of the Tests: The Vulnerable Terrorist

The effectiveness of the tests can best be seen when they are applied to situations involving the PDI. As we noted at the outset of the chapter, such situations are not nearly as prevalent as those involving application of the PDE. Two, however, come to mind. The first, regarding the question of the morality of nuclear deterrence, we shall reserve for the next chapter; the second involves the question of the limits of state-sponsored coercion in combatting terrorism.

Imagine that a terrorist has planted an automatic explosive device somewhere in the centre of London. The bomb is of such force that it will surely annihilate at least a million innocent inhabitants of the city if it is not disarmed. The police have succeeded in capturing the terrorist, but he has so far refused to divulge the bomb's location. However, the police have discovered that the man is completely and utterly devoted to his five year old daughter. They have brought her to the terrorist, and have told him that they intend to kill her unless he reveals the location of the bomb. Imagine further that a bluff by the police (at least a 'corporate bluff') would be immediately detected by the terrorist, and render the entire endeavour ineffective.[18] Are the police morally justified in making this threat?

A standard moral analysis of this situation would likely split along deontological-consequentialist lines. Deontologists would claim that the police are not justified, since their action constitutes a blatant violation of the daughter's rights. Strict consequentialists would see the police action as justified, balancing a threatened loss of one life against the certain loss of many more. Neither of these positions is wholly satisfactory, mostly because the reasoning behind each position fails to illuminate

[18] Although denying the possibility of a corporate bluff strains the credibility of the example, it is required in order to make it completely applicable to the PDI.

important issues in the case. Application of the Principle of Double Intention may shed more light on the situation.

What we must determine then is whether the example satisfies the conditions of the PDI. As to neutrality, the endeavour could be characterised as 'acting to prevent the loss of a million innocent lives.' Such an enterprise is surely at least morally neutral, so the first condition is satisfied.

We pass on to the determination of the primary intention. In this instance the two intentions of moral concern are the good intention to prevent a London holocaust and the evil intention to kill the terrorist's daughter. At this point we may use the qualification tests to reveal the primary intention. However, before discussing those tests I shall complete the discussion of the remaining conditions.

The third condition prohibits execution of the evil intention. Is killing the daughter a required part of the plan to save the lives of the million innocents? Clearly not. Indeed, actually carrying out the daughter's murder signifies that the project is lost; it certainly serves no conceivable purpose in the endeavour. As was discussed earlier (§3.3), this of course does not imply that the intention is merely a bluff.

The final condition of proportionality will be satisfied if one can verify a favourable balance of good (or rather of less bad) in the harm of forming the intention to murder the daughter over the harm of not endeavouring to save the million lives. If this is in fact a correct description of the relevant balance of negative utilities, the scales seem to tip decisively in favour of endeavouring, since merely forming an intention produces little harm in comparison with the massive loss of innocent life, even if we include some factor for the increased risk that the daughter will actually be killed. Most utilitarians would have no difficulty in approving the police tactics.

Let us return then to the second condition, and the qualification tests. Applying the Countermeasures Test to the case, we get: Do the police adopt all reasonable

means to ensure that the murder of the daughter does not occur? While details available in this hypothetical example are incomplete, there is no reason to suppose that the police are negligent in their treatment of the daughter, at least with respect to her intended death. It might strongly be argued that they have seriously violated her rights, but not in a way which would fail the Countermeasures Test.

The Nonrequisite Test can be instantiated in this case as: Would the police still have developed and maintained the intention to murder the daughter if, contrary to their expectations, that intention were not required for the success of the endeavour? The answer here must be no. The police objective is to prevent the loss of many lives; the death of the daughter, or rather the intention to kill the daughter, can only be viewed within that context, and not as an independent event. Certainly it is reasonable to assume that if the police had discovered another means to locate and diffuse the bomb, they would not have formed (or maintained) the evil intention; there would be no point in doing so. Indeed, if there were an independent reason for so intending, they would have failed to qualify for justification under the PDI.

The conditions having thus been satisfied, the police action is justified under the principle. Forming the intention to kill the daughter can be permitted.

5. IMPORTANCE OF THE PRINCIPLE

Admittedly, the PDI has limited applicability. Intentions within a given endeavour are normally rather homogenous with respect to their moral status; it is rare to find diametrically opposed intentions struggling within one enterprise. But the uniqueness of those situations in which one does find such opposition makes them just that much more difficult to assess. The difficulty of such cases can be eased by reference to a principle whose purpose is to ferret out acceptable from unacceptable endeavours.

Designed specifically to accomplish that task, the PDI recoups in importance what it gives away in applicability. As I mentioned above, the principle can be applied to policies of nuclear deterrence, which contain intentions that cannot be readily reconciled by appeal to other moral principles. That alone makes the principle worthy of consideration. It is that application which will be the central focus of the next chapter.

10: APPLYING THE PRINCIPLE TO NUCLEAR DETERRENCE

---

1. DEFINING DETERRENCE 171

2. APPLYING THE PRINCIPLE OF DOUBLE INTENTION 173

   2.1 The Conditions Satisfied 173
   2.2 Primacy of the Good Intention 176
   2.2.1 The Countermeasures Test
   2.2.2 The Nonrequisite Test

3. 'THE PRINCIPLE IS *AD HOC*' 179

4. FURTHER IMPLICATIONS 181

10: APPLYING THE PRINCIPLE TO NUCLEAR DETERRENCE

We seem to be up against a plausibly irresolvable problem in the notion of an intention in contexts of complex strategic interaction.

-Russell Hardin

Having set out and defended the Principle of Double Intention, we are at last in a position to apply that principle to the problem at hand, namely a typical policy of nuclear deterrence.[1] This application should settle the question of whether the deterring agent (nation) is justified in forming the conditional intention to retaliate.

1. DEFINING DETERRENCE

In Chapter 2, we defined deterrence by stipulating that *one deters when one endeavours to prevent another from achieving a particular goal by developing a barrier to achievement of that goal which is recognized as credible,* where an endeavour is a

---

[1] Although our focus has been on a nuclear deterrence policy as exemplified by the countries of NATO, it is my belief that such a policy is virtually identical to that of the Warsaw Pact allies (at least in those respects most relevant to this thesis), so that conclusions about one will apply equally to the other. This is particularly true with respect to the stated means and intentions of both policies: William Lee's conclusion about Soviet deterrent intentions (p. 91) is that 'the Soviets do not consider populations and cities valid targets.' This compares favourably with William Clark's letter to the US Catholic Bishops (para. 179, n. 81), in which he states that 'the United States does not target the Soviet population as such.' The similarity seems also to hold true with respect to overall deterrence policy objectives.

complex policy of actions, intentions, and intermediate results (e.g., procurement and deployment of weapons systems, development of response plans, etc.) aimed at achieving a single goal.

For the policy of nuclear deterrence, that primary goal is peace and security, accomplished by deterring enemy aggression against one's nation or one's 'vital interests'. It is important, although perhaps obvious, to note that the purpose of nuclear deterrence is not (simply) the prevention of nuclear war. Punitive deterrence is not the most effective means of preventing war; unconditional capitulation, for example, would be much more efficient. But it would also be unacceptable: Deterring western governments seek to defend values as well as prevent war. Both of these objectives can be captured under the heading of 'deterrent prevention of aggression'.

This goal is achieved by developing a credible potential threat which is intended to 'make aggression an unacceptable option.'[2] The threat is made credible by what I have called the hardware and software of deterrence, that is, the nuclear warheads and support equipment, plus the training and preparation of the military personnel whose job it would be to carry out the retaliation order.

Deterrence by threat of retaliation is at present the only available means to achieve the desired prevention of aggression. Other methods are being explored (e.g., the United States' Strategic Defence Initiative), but they are as yet mere possibilities, not feasible in the near future.

The entire enterprise of nuclear deterrence is motivated by the intention to act so as to deter aggression, what I have labeled I_p. This is the driving force behind the conglomeration of intentions, acts and results which make up the enterprise. But at the heart of the necessary threat lies the secondary (conditional) intention to retaliate,

---

[2]From the British Government's 'Replies of Foreign and Commonwealth Office,' April, 1983, as quoted in Finnis, et al., p. 8.

I_S. It may be argued that I_S is an integral part of the 'software' of preparation and training, although the possibility of an atomistic bluff (i.e., a deterrent without I_S) casts some doubt on this claim. In any case, we may accept that I_S is necessary for the achievement of the goal of deterrence.

And therein lies the moral problem. Both intentions are necessary for success. The first, I_P, is not terribly perplexing. Indeed, the intention to preserve peace and security is laudatory. However, we have seen that the second, I_S, is intrinsically immoral. The question we must now answer is this: Is the agent who undertakes the endeavour of deterrence justified in forming this immoral secondary intention?

2. APPLYING THE PRINCIPLE OF DOUBLE INTENTION

We can answer the question of agent justification by appealing to the Principle of Double Intention, which was specifically developed to deal with agent morality in situations involving evil intentions.

That the PDI seems to be the correct vehicle for assessing nuclear deterrence is evident from the fact that deterrence is an endeavour with the appropriate kinds of morally opposite intentions, a good intention, I_P, which embodies the goal of the endeavour, and an intrinsically evil intention, I_S, which is judged to be evil because it is directed on the admittedly immoral act of retaliation. Does deterrence satisfy the conditions of the PDI?

2.1 The Conditions Satisfied

The first condition of the PDI requires that the endeavour be morally acceptable. The endeavour of deterrence, which can be identified by its singular goal, is 'achieve-

ment of peace and security through deterrent prevention of aggression'. Although the determination of moral acceptability is the proper function of the encompassing moral theory, and not of the principle itself, it seems quite certain that (regardless of which theory is employed) the goal of deterrence is at least morally neutral (the minimum requirement for this condition), and arguably is one of most morally praiseworthy goals for a nation to seek. Thus the first condition is satisfied.

Skipping to the third condition, we seek to determine if execution of the evil intention is an integral part of the endeavour. In the case in question, the answer is no. Launch of a retaliatory strike is not required for the success of the endeavour of nuclear deterrence, for retaliation would be ordered only if that endeavour failed. A launch cannot further the agent's cause; it is purposeless.[3] Indeed, with respect to deterrence, it is counterproductive. A nation facing imminent nuclear catastrophe has no rational incentive to escalate the damage by launching a retaliatory attack against what is likely to be the only nation which can possibly help it with any type of recovery program. So even though the intention to launch is necessary, the execution of that intention is not envisioned, desired or required. It is simply not a part of the enterprise. The third condition is satisfied.

The fourth condition asks if there is an acceptable balance of (negative) utility when comparing the harm of not endeavouring with that of intending the evil act. In this case, that equation amounts to the harm of not acting to deter enemy aggression compared with the harm of forming and maintaining the intention to retaliate. The former harm may be as difficult to assess as the latter. There is a vast difference of opinion on how much negative utility will actually be realised by a NATO decision not to act against Soviet aggression. Anthony Kenny argues that abandoning western nuclear defences does not guarantee that the Soviets will begin a campaign of world

---

[3] See Schelling, p. 187: 'The purpose is deterrence *ex ante*, not revenge *ex post*.'

domination, and that even if they did so, ours would certainly not be a fate worse than death.[4] Alternatively, Finnis & co. argue that unopposed Soviet aggression would soon lead to the loss in the West of those precious values and freedoms which are the instruments of our greatest happiness. Such a loss would constitute an enormous harm of global proportions.[5] While negative utility is a relative quantity, we may agree to a certain extent with the Finnis view that at least some negative utility will be realised by abandoning deterrence. And because of the deep fundamental differences between western and Soviet ideologies, this will continue to be true, at least until the impressive reforms begin to alter the foundational underpinnings of the Soviet political system, as they have already begun to do in many of the Warsaw Pact nations.

We must weigh this harm against that resulting from forming the evil intention, I_S. As I mentioned in Chapter 9, this is difficult to quantify and compare, but we can consider it to be the composite of the direct harm caused by intending to retaliate, plus the indirect harm of the increased risk of retaliation. The direct harm of intending will probably be in the form of negative utility for the agent himself: We might say of that agent that he has seriously damaged his moral character, that he is, to use Kenny's words, 'a man with murder in his heart.'[6] But the problem with this characterisation is that it begs the question of whether the agent is morally justified in forming the intention to retaliate. Character damage is precisely what we are trying to determine by applying the PDI. The best that can be said about the agent at this point is that there is *prima facie* evidence of harm to his character.

The indirect harm is a factor of the additional risk of actually performing the evil incurred by forming the intention to retaliate. Developing the intention certainly must

---

[4] See Kenny, 1984, pp. 12-27, and 1985, pp. 34-36.

[5] Finnis, et al., pp. 70-74.

[6] Kenny, 1985, p. 56.

increase the risk that retaliation will take place, and so that risk should be considered. But before the harm of that risk can be entered into any utility calculation, it must be balanced against the benefit of forming the intention, which in turn arises from the autonomous effects of that intention which increase the likelihood that the deterrent will be successful. However, the actual utility calculation is an empirical matter and thus we must at this stage remain neutral about whether or not the proportionality condition is satisfied.

2.2 Primacy of the Good Intention

An acceptable determination of the second condition requires more than a superficial glance at the endeavour of deterrence, especially since, as I have designated the intentions, it would be sophistical to quickly claim that I_p is the primary intention. A bit more proof is needed. That proof will be in the form of the two qualification tests for the PDI. These tests should settle the issue of the agent's fundamental attitude toward the evil intention, an issue which is vital to an assessment of his moral worth.

2.2.1 The Countermeasures Test

The Countermeasures Test is used to examine the preventive measures the agent has taken to ensure against execution of the evil intention. It may be instantiated:

> Does the deterring agent adopt all reasonable means to ensure that the intention to retaliate is not executed?

On the whole, the answer is yes. A significant portion of defence resources allocated to deterrence are expended on hardware and software methods designed to guarantee

against both accidental and unauthorised use of nuclear warheads, as well as execution based on erroneous information. Unlike many conventional weapons systems where a decentralised command and control system is essential for their effective use, strategic nuclear weapons systems are designed to be pyramidally centralised in order to prevent use, rather than expedite it, a fact which also points to their unique role as preventers, not arresters, of aggression.[7] Additionally, should deterrence fail, many employment scenarios are designed to terminate nuclear weapon exchange at the earliest possible moment and at the lowest level of escalation, thereby further reducing the dangers of global catastrophe.[8] It should also be noted that deterring nations in both superpower blocs have recently increased their reliance on confidence-building measures designed to decrease world tensions, and therefore the chances of nuclear war. Such steps must be considered as countermeasures to execution.[9]

Against satisfaction of this condition, it must be said that some types of nuclear weapons (in particular immobile land-based ballistic missiles and strategic bomber aircraft) are inherently more vulnerable to attack than others (e.g., submarine-based missiles). This increased vulnerability leads to a shortened decision time for execution, and that in turn increases the risk of an erroneous launch due to faulty information about an enemy's actions. It seems then that not every reasonable countermeasure is being employed.

Although it is questionable whether this failure to take advantage of every

---

[7] As we noted in Chapter 4 (§2.1), Finnis & co. (pp. 56-58) argue that the command and control structure of the deterrence apparatus is unstable, and especially vulnerable to inadvertent execution following a decapitation strike aimed at eliminating the upper echelons of command. They claim that in certain scenarios, launch of nuclear weapons is designed to occur unless specific countermanding orders are received, which of course would not be forthcoming after such a strike. However their claim that the classified control structure is designed to 'fail-deadly' is unsubstantiated. And even if the claim were true, the danger of decapitation is low, given that the attacking nation would realise that this type of attack virtually eliminates any chance of a negotiated settlement of the conflict, something which must be a priority for any warring nation.

[8] The existence of the scenarios also casts doubt on the acceptability of the escalation hypothesis discussed in Chapter 2 (§2.3).

[9] For a detailed account of the types and effectiveness of confidence-building measures, see e.g., Alford, esp. pp. 5-8.

available option seriously damages nuclear deterrence *vis-à-vis* the PDI, it certainly points out that there may be some forms of deterrence which are more satisfactory than others. An ideal deterrent force would probably eliminate land- and air-based weapons altogether, in favour of a virtually invulnerable (albeit somewhat less accurate) submarine-based force.[10]

2.2.2 The Nonrequisite Test

The Nonrequisite Test is the final determinant of the agent's fundamental attitude toward the evil intention. In this case it asks:

> Would the deterring agent still develop the intention to retaliate if, contrary to his beliefs and expectations, it were not required to prevent aggression?

The answer here is no. The intention itself holds no intrinsic value for the agent; it is valuable only as an instrument to achieve the objective of deterrence. Although at present beliefs and expectations point to the formation of I_s as the only way to succeed, the fact that at least one western power is examining alternatives (SDI) indicates that there is no firm attachment to the evil intention.

That there is no independent reason to maintain the intention is also shown by the fact that I_s is formed only with an eye toward potential aggressors with a *reciprocal nuclear capability*. Great Britain, for example, did not develop (or at least announce) a conditional intention to launch against Argentina if the attempt to recover the Falklands proved to be unsuccessful. Neither did the United States in its campaigns against Libya, nor the Soviet Union during its conflict in Afghanistan. Significantly, the United States under President Truman made no (genuine) effort to announce

---

[10] This concept of force reduction plays a key role in Kenny's proposal for disarmament (1985, pp. 70-71), although he argues only for the temporary retention of an SLBM force. I shall return to the issue in Chapter 12.

an intention to use nuclear weapons against Japan, a requirement for authentic deterrence. The difference was that the early use of nuclear weapons was for winning war, rather than deterring aggression. Recent efforts to reduce the superpower arsenals also bespeak of the solely instrumental value of the retaliatory intention. Reductions in arms would run counter to a preference for the intention to inflict massive damage.

In our examination of bluffing in Chapter 4, we saw that there may be options which would involve abandoning the secondary intention. It would of course be incumbent upon any deterring nation seeking moral justification under the PDI to explore the viability of these options.

Having thus satisfied the qualification tests as well as the four conditions of the PDI, the agent (nation) who undertakes a policy of nuclear deterrence is justified in forming and maintaining the conditional intention to retaliate, despite its immorality.

3. 'THE PRINCIPLE IS *AD HOC*'

At this point, one may object that the PDI is nothing more than an *ad hoc* hypothesis sophistically designed to rescue the obviously immoral policy of punitive nuclear deterrence. It is a 'designer theory' specially formulated to support the otherwise questionable contention that forming I_s, and therefore developing a policy of nuclear deterrence, is morally justifiable. The PDI has no other purpose, and little to recommend itself as a *bona fide* moral principle. It is a mistake to develop principles for which there is little or no independent support simply to rescue a further unsupported claim.

Before answering this charge, we should try to clarify exactly what it is that differentiates a genuine principle of moral theory from one which should be rejected as *ad hoc*. But as we try to do so, it becomes clear is that the differentiation itself is

a matter of dispute. As J. L. Mackie notes, it is difficult to 'systematically mark off this error from the respectable procedure of interpreting new observations in the light of an established theory.'[11] Certainly a principle which is blatantly counter-intuitive (e.g., a superstition) must be considered *ad hoc*. Equally certainly one cannot condemn a principle simply on the basis of either its origins or applicability without also committing a genetic fallacy. But between these two extremes the graduated scale of acceptability is not well calibrated. There seem to be no established criteria for deciding what constitutes a sound principle, except perhaps the relative ability to account for the observed data.

On the basis of this last ability, the PDI must survive the charge against it. It was developed to help explain the observed impression that nuclear deterrence is the necessary means to an end the attainment of which may well be obligatory, despite the fact that the endeavour incorporates contradictory moral intentions.[12] And although the principle was put forth to reconcile the good and the evil of nuclear deterrence, its application extends beyond that limited field to include many other types of deterrent endeavours. As we have seen, the principle can account for a state's right to punish as well as a parent's permission to discipline by using a threatened sanction. In general, it can explain why an isolated intention to perform a wrong act may be justified within a larger context. This general applicability, when coupled with its demonstrable plausibility, is more than sufficient to refute the charge that the PDI is merely an *ad hoc* attempt to rescue nuclear deterrence from moral condemnation.

---

[11]Mackie, p. 175.

[12]See Finnis et al., pp. 77-78, where they accept that there is a 'grave obligation' for the West to oppose Soviet power.

4. FURTHER IMPLICATIONS

With that we come to the end of the moral analysis of nuclear deterrence. The embedded conditional intention to retaliate is granted to be *prima facie* immoral. Without strong evidence to the contrary, this judgement will lead to the condemnation of an agent who forms such an intention. But appeal to the Principle of Double Intention provides that necessary evidence. It demonstrates that there are acceptable reasons for forming the evil intention, and this in turn provides a moral justification for the deterring agent.

But we have yet to answer the earlier charge that forming the intention to retaliate is not rationally possible for an agent who acknowledges the irrationality of retaliation itself. However, armed with a moral justification for the intention, we are now in a position to confront the rationality problem.

11: AGENT RATIONALITY AND THE SECONDARY INTENTION

---

1. CONFRONTING THE QUESTION OF RATIONALITY 183

   1.1 Significance of the Question 184

2. THE PROBLEM OF AGENT RATIONALITY 185

   2.1 *Akrasia* and Rationality 186
   2.2 The Toxin Puzzle 187

3. AGENT RATIONALITY SOLVED 190

   3.1 *Akrasia* and Deterrence 190
   3.2 The Toxin Puzzle and Deterrence 193

4. THE INTENTION-BELIEF CONSISTENCY OBJECTION 197

   4.1 The Intention-Belief Argument Against Deterrence 198
   4.2 Possible Answers to the Objection 199
   4.3 Denying the Need for Positive Belief 201
   4.4 Questioning The Escalation Hypothesis 203

5. AGENT MORALITY AND DETERRENCE 204

11: AGENT RATIONALITY AND THE SECONDARY INTENTION

To this point we have developed a suitable notion of ordinary intention, and have shown deterrent intention to be a unique combination of two distinct intention components. We have examined the modal question of necessity with regard to the secondary intention, and have accepted for the moment that it must be present for effective deterrence. And we have applied the Principle of Double Intention to nuclear deterrence, and thereby found a possible moral justification for forming the secondary intention to retaliate.

We now finally turn to face the problem which has awaited us at every juncture: How is it possible, given the admitted immorality and irrationality of retaliation, for the deterring agent to form the intention to retaliate and yet retain his rationality? It seems that this cannot be done. The agent, if he is to deter successfully, must sacrifice his rationality in order to form the requisite intention.

1. CONFRONTING THE QUESTION OF RATIONALITY

I shall concentrate in this chapter on the question of the rationality of the deterring

agent, which mirrors the necessity question of Chapter 4. For what we must determine before we can morally exonerate the deterring agent is whether it is even possible for him to form a rational intention to retaliate, given that he has no reason to act on it. That is, we must determine if in general it is rationally possible for an agent to form an intention knowing that there are conclusive reasons against acting on the intention.[1]

1.1 Significance of the Question

Before attempting an answer, we should underscore the importance of finding the correct one. If forming such an intention is rationally possible, then the defender of deterrence may appeal to the PDI to justify the intention. But if it is not rationally possible to form the intention given the irrationality of acting on it, then the defender faces a much more difficult task. He would have to show that the (apparently) conclusive reasons against retaliation are not in fact conclusive. To do that, he would need to demonstrate that retaliation itself is rationally possible, perhaps by showing that the reasons against retaliating are not conclusive, but merely strong *prima facie* reasons, and only then go on to argue for the relative benefits of aggression prevention over genocide. However, neither of these routes seems to hold much promise. Indeed, we accepted in Chapter 2 that retaliation is irrational as well as immoral, given the acceptability of the escalation hypothesis.

However, a third possibility exists. The defender may accept that the reasons against *acting to* retaliate are conclusive, but argue that the reasons against *intending to* retaliate are not thereby necessarily conclusive, but may be overridable. That done, he would once again be in a position to weigh the relative benefits of aggression

---

[1] I have chosen to phrase the question in terms of whether forming the intention is 'rationally possible' in order to convey the problem as whether a rational agent can intend to do something which is irrational. That is, can he so intend and also maintain his rationality, or must he sacrifice that rationality in order to form the intention?

prevention against, not genocide *per se*, but rather the increased risk of acting on that intention.

This is an enticing possibility, but it seems that we are quickly led back to the original question of the rational possibility of forming an intention to act irrationally. What I shall show in this chapter is that while this may in general remain a difficult puzzle, it does not apply to nuclear deterrence because of the dualistic nature of deterrent intention and the doubt surrounding the escalation hypothesis.

2. THE PROBLEM OF AGENT RATIONALITY

To begin this discussion of the rationality of deterrence, we must settle on some acceptable notion of what it means to be rational. In this context, I shall take a 'rational act' to be one which is 'best supported by reasons,' viz., the product of Aristotelian practical syllogism, and shall then take an 'irrational act' to be one which is done 'contrary to, or without the best reasons.' An act is irrational just when an agent has no reasonable basis for choosing it from among the alternatives. A purely arbitrary, capricious decision is irrational in this way. Similarly, I shall describe an agent as 'rational' just when he possesses and exercises an ability to reason, and 'irrational' when he loses or fails to demonstrate that ability, that is, when he has no reasonable basis for the choices he makes.

While this broadly intuitive notion of rationality is admittedly superficial, and thus ignores many important issues inherent in a complete definition, e.g., what constitutes an acceptable reason, the importance of the agent's attitude toward discovering an appropriate reason for action, etc., we shall see that it is more than sufficient to impugn the type of defence of nuclear deterrence which we are analysing.

2.1 *Akrasia* and Rationality

Before introducing the rationality question with regard to the relatively complicated case of acting with a further intention (of which deterrence is an instance), we may consider first the simpler problem which we encountered in Chapter 3 concerning intentional action and *akrasia*. Although we considered the problem of *akrasia* only as a possible objection to the linear relationship model, it also has a bearing on the present question of agent rationality. Consider for example Davidson's definition of *akrasia:*

> In doing x an agent acts incontinently if and only if: (a) the agent does x intentionally; (b) the agent believes there is an alternative action y open to him; and (c) the agent judges that, all things considered, it would be better to do y than to do x.[2]

Condition (c) implies that y is the result of practical reasoning, and so the agent's act of doing x seems to be irrational, in the sense that it is done without (the best) reason in mind. And by implication, the agent must also be irrational, since he has failed to demonstrate his ability to reason correctly, where such demonstration would take the form of acting on the basis of his deliberation.

This much at least is acceptable: It seems irrational for an agent to conclude that he ought not do x (viz., that he believes that there are convincing reasons against doing x when compared to doing y), but nevertheless to do x anyway. To carry out a process of practical reasoning and arrive at a decision to act, but then to abandon that decision is, given our abbreviated definition above, irrational. This view, which seems correct, gives rise to the further claim that one cannot rationally *intend* to act against the conclusions of one's practical reasoning, viz., that it is irrational to form the intention to act irrationally.

---

[2]Davidson, p. 22.

But this line of thought does not help us to resolve the present question of agent rationality in deterrence. First of all, the last claim about intention does not follow from the claim in the simpler case that one's actions against the conclusions of practical reasoning, although intentional, are irrational. Imagine for instance that an agent concludes against *x-ing*, but nevertheless considers forming an intention to *x*. The claim is that he can do so only by sacrificing his rationality. But Davidson's claim that one acts incontinently when one does that which he has decided against doing does not imply that it is also irrational to form that intention to so act, unless one can prove an absolute version of something like the wrongful intentions principle for rationality, e.g., that it is always irrational to intend to do that which is irrational to do. However, as with the wrongful intentions principle itself, the strongest supportable version would be a defeasible principle that it is *prima facie* irrational to intend irrational acts. Thus the claim that one cannot rationally intend to act against the conclusions of one's practical reasoning does not immediately follow from Davidson's definition. Nor is it self-evidently true.

And even if the claim were true, it has little relevance to the rationality objection against deterrence. As Anscombe has pointed out, there is a conceptual gap between acting intentionally and the expression of an intention.[3] The problem of irrational action is distanced from deterrence, since deterrence involves only intention formation, and not intentional action directly.

2.2 The Toxin Puzzle

Before examining a more formal defence along these lines, we should first consider a more difficult (and perhaps more relevant) problem. The question of the rational

---

[3] Anscombe, 1966, §1.

possibility of deterrence may be viewed as a version of the Toxin Puzzle broached by Gregory Kavka and analysed by Michael Bratman.[4] The puzzle is this: You will be paid one million pounds tomorrow morning if at midnight tonight you *intend* to drink a bottle of toxin tomorrow afternoon. The toxin will only make you very sick for a day, and will have no after-effects. And in fact, the money will be yours if you simply form the intention; you do not need to actually drink the toxin. The problem is of course that you have great incentive to form the intention, but no incentive (and an important dis-incentive) to actually carry out your intention. But one cannot form an intention to perform that which one has no reason to perform. And so you will lose your chance to become a millionaire.

According to Kavka, the puzzle arises because intentions are 'dispositions to act based on *reasons to act*--features of the act itself or its (possible) consequences that are valued by the agent.'[5] Thus they can only be formed with those reasons in mind. Without reference to those reasons, there can be no intention.

The puzzle turns on the very close connection between the rationality of forming an intention and the rationality of acting on that intention. Bratman, in an attempt to answer the objections to his action theory which the puzzle raises, analyses the argument underlying the toxin problem, and tries to show that it results from a confusion between present-directed and future-directed intentions, viz., intentions to act at some time in the future.[6] However, to see the problem more clearly in light of the present question, I have adapted Bratman's argument, and applied it to nuclear

---

[4] Kavka, 1983, pp. 33-36; Bratman, pp. 101-106.

[5] Kavka, 1983, p. 35.

[6] Bratman, pp. 102-06.

deterrence:[7]

(1) It is rational of the deterring agent at t_1 to (conditionally) intend to retaliate at t_2 (where t_2 produces the relevant retaliation conditions), given his strong reasons at t_1 for forming that intention.

(2) If it is rational of the deterring agent at t_1 to intend to retaliate at t_2, and if t_2 does produce the relevant retaliation conditions, then it is rational of the agent not to reconsider his intention then.

(3) At t_2, the relevant conditions do occur, and therefore the agent does not reconsider his prior intention.

(4) If it is rational for an agent to have a present-directed intention to *x*, and he successfully executes this intention and thereby intentionally *x*'s, then it is rational of him to *x*.

(5) It is rational for an agent to maintain a future-directed intention to *x* at t_2 just in case (a) it was rational originally to form this intention, and (b) it was rational of that agent from t_1 to now not to reconsider the intention.

(6) Therefore, it is rational at t_2 for the agent to retaliate.

But of course this conclusion contradicts our earlier assumption that it would be irrational of the deterring agent to retaliate. So at least one of the premises (1)-(4) must be rejected. Bratman argues that we should abandon (1), since 'in deliberation about the future we deliberate about what to *do* then, not what to intend now.'[8] That is, the conclusion of the agent's practical reasoning is a choice to act (in the future), not simply a choice to intend. Therefore, reasons which will influence intention formation alone (without reference to the act intended) cannot affect deliberation. Intention cannot be distanced from action in the way which premise (1) requires.

The idea that intentions are not the end product of practical reasoning, but only

---

[7] In concluding his article about the puzzle, Kavka (p. 36) notes its similarity to the paradoxes of deterrence which he discussed in an earlier article (and which we analysed in Chapter 5). The difference with the Toxin Puzzle, he says, is that it concerns unconditional intentions, and therefore broadens the application of the earlier discussion. Although it may seem that I am once more limiting the scope of the problem by refocusing on deterrence, it will become clear that my solution to the puzzle, while specific to deterrent intentions, does not hinge on their conditionality.

[8] Bratman, p. 103.

an intervening step on the way to action, seems generally right and certainly explains the attraction of moral principles such as the wrongful intentions principle which seek to bind together acts and intentions. However, this answer does not solve the puzzle, but merely sidesteps it. The puzzle remains, especially for deterrence. We cannot surmount the rationality hurdle and rationally intend an irrational act; we cannot recognise conclusive reasons against acting and yet form the intention to act.

3. AGENT RATIONALITY SOLVED

The problems posed by *akrasia* and the toxin puzzle combine to confront the deterring agent with the serious charge of irrationality. The rational impossibility of acting against the dictates of practical reason on one hand and intending to act without reason on the other pose a significant threat to the credibility of the dualistic interpretation of deterrent intention which we have set forth. But the charge can be answered. And in both cases, the answer comes in the form of accepting the verity of the two problems, but denying their applicability to deterrence.

3.1 *Akrasia* and Deterrence

There are two ways to show that the problem of acting contrary to one's best judgement does not affect the case of deterrence. The first is to recall that Davidson's conditions of *akrasia* address only the problem of akratic action; they do not readily apply to akratic intentions. At issue in deterrence is whether it is rationally possible to intend the irrational act of retaliation, not whether it is rationally possible to actually carry out that retaliation. It may be true that an agent who in fact retaliates does so knowing that there is a better act (e.g., *not* retaliating) open to him, and thus acts

akratically according to Davidson's criteria. However, this does not transfer to a judgement about the irrationality of intending *per se* unless one can prove, against our conclusions in Chapter 3, that there is an intrinsic, indivisible connection between act and intention in general, and between the secondary intention of deterrence and the act of retaliation in particular.

The second way of showing the inapplicability of the *akrasia* problem to deterrence is to recall that the deterring agent does not form the secondary intention in isolation, nor does he form it out of any preference for retaliation. Rather, he forms it within a larger context whose importance cannot be overstated. This point is directly relevant to Christopher Peacocke's claim about *akrasia:*

> The *akrates* is irrational because although he intentionally does something for which he has some reason, there is a wider set of reasons he has relative to which he does not judge what he does to be rational.[9]

The motivation behind this statement is the problem of explaining akratic action which is both intentional and irrational. Generally, to act intentionally is to act for a reason, but to act irrationally is to act without a reason. How then can akratic action be both intentional and irrational? Peacocke's answer about a wider set of reasons implicitly distinguishes 'acting for a reason' from 'acting for the best reason.' Within the narrow set of reasons for acting there may be one which provides justification for doing *x*. It may be, for example, that I decide to smoke because it relaxes me. But the wider set of reasons may contain justification for not doing *x* which supersedes the narrower reason. Within that wider set may be my realisation that smoking is extremely hazardous to my health, a reason which outweighs the benefits of relaxation. Knowing this, my smoking is intentional relative to the narrow set and irrational relative to the wider set, and thus akratic.

---

[9] Peacocke, p. 52

Taking the cue from Peacocke's distinction, and given our conclusion in Chapter 3 that intentions are a proper subject of independent rational as well as moral analysis, we can see a solution for deterrence: Examined in isolation, the intention to retaliate is irrational. That is, there is no reason to intend to act irrationally. But viewed within the wider context of the endeavour of deterrence, it is not irrational to form that intention. For that context includes a set of reasons relative to which the deterring agent judges his intention formation to be morally justified (by the Principle of Double Intention) and therefore rational, i.e., reason-based. Unlike the narrow context, which only includes reasons for intending to retaliate, the wider context of the dualistic interpretation of deterrent intention also includes reasons for forming the primary intention to deter aggression.

The dualistic model also begins to answer D. F. Pears' potential objection which goes to the heart of the rationality problem. Pears argues that if an agent believed that his intention would reduce its own effectiveness (i.e, its ability to lead to action) to zero, 'he simply could not form it.'[10] This certainly seems to be true for intentions taken in isolation, since it is normally the case that an agent must believe that his forming an intention increases the probability that he will perform the intended act. But it does not apply to the embedded secondary intention in deterrent situations, where the agent has a larger set of reasons based on which he prefers, and through deliberation intends, to militate against the effectiveness of his more narrow intention. That is, the deterring agent forms the narrow secondary intention within a context in which the effectiveness of that intention is *purposely* reduced to as close to nil as possible. Therefore, it is rational for him within that context to form that narrow

[10]Pears, p. 80.

intention.[11]

3.2 The Toxin Puzzle and Deterrence

But this response does not yet answer the more relevant intention-belief objection posed by the toxin puzzle, i.e., the impossibility of forming a future-directed intention to act irrationally. That answer comes in the form of distancing deterrence from the puzzle.

Before laying out that answer, we may begin by noting that there is a semantic reason why the puzzle fails to adversely affect the rationality of deterrence. Implicit in the puzzle are two senses of 'adopt' which serve to explain why an agent adopts a particular intention, senses which Bratman explicitly draws out:[12] First, an agent may adopt an intention on the basis of practical reasoning about the act. That is, he will reason with an eye toward action, settling on an intention merely enroute to that action. Secondly, an agent may adopt an intention as a non-reasoned acquisition; he may form an intention without reasoning about it at all, but rather as a result of, say, self-hypnosis or direct revelation.

Leaving aside the question posed by the second sense of 'adopt' of whether an intention can be formed irrationally (in the sense that it is formed without reason), it is clear that these two do not exhaust the list of possible ways in which one can adopt an intention. There is at least a third interpretation of 'adopt' which is critically important to deterrence: An agent may adopt an intention on the basis of deliberation about simply whether to form that intention, without reference solely to the act

---

[11]David Gauthier makes a similar point about basing rationality on the larger endeavour rather than on the individual action: 'If [one] accepts deterrent policies, then [one] cannot consistently reject the actions they require and so cannot claim that such actions should not be performed.' (p. 487). However, our respective positions differ dramatically in that he believes retaliation to be among the actions required by deterrence, and thus argues for the rationality of retaliation itself.

[12]Bratman, pp. 105-06.

intended. This would be the case if the agent had reasons to form the intention apart from reasons either for or against acting, such as we have seen to be the case for nuclear deterrence. The reasons for forming I_s have to do with the beneficial deterrent effects which that intention produces, reasons which are unconnected with the act of retaliation itself. Overlooking this third sense of 'adopt' naturally leads to an acceptance of the intrinsic, indivisible bond between act and intention implicit in the toxin puzzle. But the third version of 'adopt', supported by the dualistic analysis of deterrence, calls into question that bond.

It may be objected that introducing an interpretation of 'adopt' which relies on allowing the formation of intentions without sole reference to acts begs the very question of rationality which we are seeking to resolve. To simply assert that the conclusion of practical reasoning can be intention formation implicitly denies everything which we have accepted about practical reasoning, where the outcome is an action, not an intention. Furthermore, this interpretation of 'adopt' cannot support a complete isolation of intention formation from action. Even though it might be true that an agent has reasons to intend which are not reasons to act, that does not imply that he can adopt the intention without also considering the action itself.

To answer this second objection first, the third sense of 'adopt' is not offered to show that intention formation is possible without any reference to action. It is suggested only to show that an intention itself may be preferable for reasons in addition to those supporting the decision to act. It is meant to show, for instance, that even though an act may be absolutely forbidden, the intention to perform that act may itself be only *prima facie* wrong, given the independent justification of its formation.

And even if this interpretation of 'adopt' begs the question of rationality for the general issue of intending irrational acts, it does not do so in the specific instance of nuclear deterrence. We have seen that within that endeavour, the secondary

intention plays a unique dual role, and is therefore valued by the deterring agent apart from the normal worth of an ordinary intention. The general situation, where intentions have no value for an agent apart from their connection to the intended act or its consequences,[13] viz., where intentions are simply a conduit through which an agent focuses his energy to act, does not apply to the secondary intention in deterrence, which has *act-like* value for the deterring agent. And while it is true that the result of practical reasoning is action and not simply intention, the result of practical reasoning in the case of deterrence is I_s, an intention which functions as a pseudo-action flowing from the primary intention.

Recalling that unique role of I_s leads us to an acceptable resolution of the toxin puzzle: While the act of retaliation may well be irrational, forming the intention to do so is not, because the reasons for forming the secondary intention (*qua* action) are unrelated to any reasons (which there are none) for executing it. What leads the deterring agent to form I_s are the steps of practical reasoning which lead him to form the primary intention, viz., his preference to deter aggression coupled with his belief that retaliatory deterrence is the only way to fulfill that preference. I_s is merely the 'act' which concludes that reasoning process.

More importantly, there is a significant difference between the situation which is faced by the potential millionaire and the agent who contemplates engaging in deterrence. The toxin puzzle hinges on accepting a very close connection between intending to act and believing that one will do so. In some (perhaps most) situations, this connection is justified. Nevertheless, many critics question its assumption. Peacocke for example denies it, citing many cases where an agent will form an intention

---

[13]Kavka, 1983, p. 35.

knowing that his probability of success is very low.[14] I have no difficulty forming an intention to sink a 30-foot putt, despite my firm (and well-founded) belief that I shall very likely fail to do so.

The assumed connection between intention and belief is further refuted by the case of deterrence. For not only does the deterring agent not believe (in the strong sense necessary to satisfy the toxin puzzle) that he will carry out his intended act, he actually believes just the opposite. He believes that forming the intention to retaliate will serve to *prevent* him from being forced to retaliate. Indeed, his belief that he will not have to carry out his intention to retaliate is the very belief which motivates him to form that intention.

It is this lack of an intention-belief connection which distinguishes deterrence from the toxin puzzle. Even if we accept Pears' claim that an agent must possess a 'minimal future factual belief' that his intention will increase the probability (i.e., risk) of his performing the intended act to something greater than zero,[15] it is clear that the toxin puzzle does not directly affect the question of agent rationality in deterrence. The increased risk of intending is mitigated by a factor absent in the toxin case, namely, the agent's belief that the intention formation also increases the likelihood that he will not act. While intention formation must be considered the last step in preparation for action for the potential millionaire, it is not so for the deterring agent.

This difference between the two cases also serves to demonstrate the rationality of the deterring agent. For if, because of the likely deterrent effect on his potential enemies, he believes that his pseudo-action (forming I_S) reduces the probability of

---

[14] See Peacocke, p. 69, where he denies that 'if an agent intends to do something, he believes he will do it.' See also Davidson, p. 95: 'We do not necessarily believe that we will do what we intend to do . . . [since] reasons for intending to do something are in general quite different from reasons for believing one will do it.'

[15] Pears, p. 78.

his performing an irrational act (retaliating) when compared with not performing that (pseudo-) action, then he would be more rational to form the intention than to refrain. And since he does believe that, his act of intention formation (despite the fact that it is an intention to perform an irrational act) is rational, and therefore possible.

4. THE INTENTION-BELIEF CONSISTENCY OBJECTION

But this line of argument represents only the start of an adequate defence of deterrence against the rationality problem which centres on the close connection between intention and belief. One may well object that the discussion to this point has avoided a direct confrontation with the most serious problem posed by the rationality question. We have assumed that it is possible to separate intention and belief enough to allow that the deterring agent may rationally form the intention to retaliate without also maintaining the belief that he will do so should the situation arise. This necessary separation seems to require not merely that the agent believes the probability of his acting is very low, but that he believes it is zero. It is for this reason that the example of my intending to sink a long putt is disanalogous to deterrence. An example of that sort can only provide an accurate analogy if I believe that my chances of sinking the putt are nil, or rather, if I believe that given the opportunity to address the ball, I shall not even try to putt. For it seems that this is, after all, what it means for a rational agent to recognise 'conclusive reasons' against acting. Based on the conclusion of his reasoning, he believes that he will not attempt to act if given the opportunity. As a result, the argument purporting to show that the deterring agent can form the intention in the face of conclusive reasons against acting is faulty.

4.1 The Intention-Belief Argument Against Deterrence

Or so it appears. However, this objection moves a bit quickly to its conclusion. In order to determine its potency we must analyse it more carefully. The objection proceeds from the claim that intentions are accompanied by some minimal performance belief to the conclusion that forming the intention to retaliate is impossible. In its strongest form the argument would look like this:

(1) An intention is genuine if and only if it is accompanied by a concomitant belief by the agent that the probability of his performing the act intended is greater than zero.

(2) So if the intention to retaliate is genuine, then the deterring agent must also believe that the probability of his retaliating is greater than zero.

(3) Therefore, if the deterring agent believes that the probability of his retaliating is zero, then he cannot genuinely intend to retaliate.

(4) The deterring agent has conclusive reasons against retaliating.

(5) The deterring agent is fully rational (viz., he does not act on emotions or passions).

(6) If an agent has conclusive reasons against doing x and is fully rational, then he must believe that the probability of his doing x is zero.

(7) Therefore, the deterring agent must believe that the probability of his retaliating is zero.

(8) Thus by (3) and (7), the deterring agent cannot genuinely intend to retaliate.

The argument, if sound, effectively refutes the attempts in Section 3 of this chapter to answer the rationality objection by driving a conceptual wedge between intending and believing. For it shows that no such wedge can be found, since in general intention cannot be divorced from belief.

Before examining the soundness of the argument, we must clarify exactly what is meant by saying in (7) that the deterring agent must believe that the probability of

his retaliating is zero. This is not the claim made by Gerard Hughes (and disputed by Finnis & co.) that the agent believes that he will never have the occasion to retaliate, viz., that the conditions of the conditional intention to retaliate will never obtain.[16] As a simple description of external events which are for the most part independent of the agent, this interpretation reduces the belief to no more than a wager by the agent that the circumstances will not arise. As such, it is consistent with forming an intention against the wager. Thus, this interpretation of (7) will not lead to the required conclusion in (8) that forming the intention to retaliate is impossible.

Instead, (7) seems to be the more damaging claim that the deterring agent believes that he will not act to retaliate, *even if the circumstances should arise.* Here the agent is not betting against an external prediction of events. He has made a *self-*prediction about his action, regardless of external events. It is this type of prediction which seems to be inconsistent with forming and maintaining the intention to retaliate.

4.2 Possible Answers to the Objection

Given this meaning of (7), can there be an effective answer to the rationality objection, or does the objection render deterrence impossible? There are at least three possible answers we may consider. First, one may attack premise (4) by denying the assumption that the 'conclusive' reasons against retaliating are in fact conclusive, that is, that they absolutely prohibit (in the rational rather than moral sense) retaliation. The most promising approach here is to claim, against (4), that the reasons against retaliation are at most strong *prima facie* reasons which may therefore be

---

[16] Hughes, p. 33; Finnis, et al., p. 124. Hughes believes that the conditions will never obtain because the intention to retaliate itself 'will in fact ensure that it [the deterrent machinery] need never be used.' Finnis & co. deny that such an intention can ever be formed since 'one cannot conditionally intend to do what one is certain one will never have occasion to use.' Bernard Williams (p. 107) offers--and Michael Dummett (p. 122) attacks--a similar argument that forming a conditional intention to act immorally would not be wrong if one was certain that the conditions would not be fulfilled.

overridden by other considerations. This is the approach taken by David Gauthier, who concludes that retaliation is rational because the intention to retaliate is rational.[17] However, this approach, in addition to denying our assumption in Chapter 2 that retaliation would be both immoral and irrational, is radically counter-intuitive, and therefore requires a stronger supporting argument than Gauthier offers. Thus it cannot (for now at least) answer the irrationality objection.

A second possible refutation would take the form of denying the assumption of perfect rationality in premise (5) by recognising the inherent human tendency occasionally to perform acts which do not wholly admit of rational explanation. Absent this premise, the argument loses its power, since the deterring agent will recognise the possibility that he may retaliate without rational justification, for instance out of anger or a need for revenge, or simply because of the 'irreducible unpredictability of events once the nuclear threshold is crossed.'[18] Awareness of this possibility must raise the agent's assessment about the probability of his acting to some value--however small--greater than zero. And this of course will allow the agent to form the requisite intention.

But while this admission of inherent irrationality may well reflect reality,[19] it certainly does not provide an answer to the objection. To the contrary, it reinforces the idea that nuclear deterrence requires, indeed thrives on, at least the appearance of irrationality. The best that can be said here is that recognition of the possibility of irrational action makes deterrence feasible, which of course will not convince those who object to deterrence because of that very irrationality.

---

[17] Gauthier, esp. p. 486.

[18] Schell, p. 207.

[19] For instance, Schell notes (p. 205) that President Nixon cultivated a 'Madman Theory' of the Presidency, 'according to which the nation's foes would bow to the President's will if they believed that he had taken leave of his senses.' Derek Parfit (*Rreasons and Persons*, pp.12-13) offers an alternative solution by suggesting a case for 'rational irrationality', viz., that there are cases it which it is rational for an agent to cause himself to act (or, *ceteris parabis*, intend) irrationally.

4.3 Denying the Need for Positive Belief

Thus the objection cuts deeply into the argument for the rational possibility of deterrence without retaliation. And the argument supporting the main claim of the objection appears well formulated. But it is not. There is one further answer to the objection. For even if we accept the above interpretation of (7), and admit that the rationality assumption of (5) is theoretically acceptable, we must question either the truth or the applicability of premise (1).

There are two ways of understanding the meaning of the claim of (1) that all genuine intentions are accompanied by a minimal performance belief. First, one may understand it straightforwardly to mean that the intending agent must hold some positive belief about his acting on the intention. That is, he must clearly recognise the possibility that he will act.

But this strict interpretation renders the statement false. For it amounts to a severe requirement not simply for consistency, but for intention-belief isomorphism, and thus precludes the very real possibility that an agent may form an intention in the absence of such a positive belief. We may call an intention without such a belief agnostic, since the agent forms it without either believing or disbelieving that he will act on it. Bratman gives some evidence that such agnostic intentions can be formed by pointing out that we may consistently intend to do something while being aware of a tendency toward absentmindedness, a tendency which may well prevent us from forming a predictive belief about our future action.[20] It is surely possible for me to intend to meet you at 4:00 next Tuesday without believing--or more importantly disbelieving--that I will in fact do so, because for instance I know that I have a tendency to forget my appointments. The existence of agnostic intentions repudiates the

---

[20] Bratman, p. 37.

requirement for the kind of isomorphism inherent in a strict interpretation of (1). Such intentions are not strictly inconsistent since they do not provide the sort of 'head-on contradiction' which we expect inconsistent statements to display.[21] Thus, under this strict interpretation, (1) is false. And this is so not merely because we live in a contingent universe where our beliefs about our future action do not always turn out to accurately reflect reality. That is, it is not false merely because we may always be prevented from doing what we intend.[22]

Given the unacceptability of the strict interpretation, it may be possible to modify the meaning of (1) to deny the rationality of simultaneously holding an intention and a non-performance belief. That is, (1) may be understood to mean that 'Genuine intentions cannot be accompanied by a belief by the agent that he will not act on the intention should the occasion to do so arise.' This interpretation of the intention-belief claim is more liberal than the first, since it allows for agnostic intentions. What it asserts is that the agent cannot consistently hold the intention to do *x* and the belief that he will not do *x*. Having an intention to do *x* implies that the agent does not also believe that he will not do *x*, but it does not imply the stronger claim that agent believes that he will do *x*. Rather than asserting the necessity of the positive belief about *x*, the liberal interpretation simply denies the possibility of the negative belief about *not-x*. Given this interpretation, we may accept premise (1) to be true.

However, this modification weakens the premise to the point that it can no longer support the argument against deterrence. Premise (4), that the deterring agent has conclusive reasons against retaliating, does not imply that the agent believes that he will not retaliate, a necessary step given the only acceptable understanding of premise (1). The agent need not hold the required negative belief, but may instead be agnostic

---

[21] Anscombe, 1966, §52.

[22] Anscombe (§52) shows that this attempt to separate intention and belief is vacuous.

about retaliation.[23] He may not believe that he will retaliate. But this does not imply that he believes that he will not do so. The conclusive reasons of premise (4) only preclude the positive belief that retaliation will occur; they do not require the stronger belief that it will not. That stronger belief can only be supported by conclusive *moral* reasons against acting, which this argument does not address.

Given then the fact that premise (1) is either false or too weak, the argument supporting the objection of intention-belief inconsistency is unsound. Thus the objection itself can be refuted.

4.4 Questioning The Escalation Hypothesis

If, despite the above arguments, there remains a nagging doubt about whether the deterring agent can rationally form the intention to retaliate, it must at this stage be attributed to acceptance of the escalation hypothesis, viz., that any wartime nuclear detonation will inevitably touch off a series of spiraling counter-attacks, ending in worldwide devastation. For absent the assumed truth of this hypothesis, it is possible to deny the truth of premise (4), not by claiming with Gauthier that any form of retaliation is rational, but by accepting that some limited use of nuclear weapons may be rational. But the denial of (4) is impossible, given the escalation hypothesis, since any retaliation must accordingly lead to all-out destruction, which of course renders even limited use grossly purposeless and irrational.

Up to this point, we have accepted the escalation hypothesis, especially since it has been only tangentially related to the arguments presented. (The Principle of Double Intention prohibited any execution of the evil intention, irrespective of the

---

[23] As we noted in Chapter 4, many strategists argue that agnosticism about retaliation may be sufficient for effective deterrence. See, e.g., Fisher, esp. p. 79; Kavka, p. 287; Kenny, 1985, p. 79; Morris, p. 481; Schelling, p. 36; and Sterba, p. 101.

outcome of that execution.) But the rationality objection to deterrence is underscored by escalation; indeed, the significance of the objection wanes without support from that hypothesis. The objection calls into question the deterring agent's ability to form the secondary intention to retaliate because execution of that intention would be irrational. The irrationality of retaliation arises directly from the claim that the deterring agent believes that such a response would lead to levels of destruction which he would have no reason to bring about. That is, it arises directly from the escalation hypothesis. Without that hypothesis, the irrationality objection is seriously weakened, resting on the claim that the deterring agent would have no reason to launch any form of retaliation. But without the assumption of the escalation hypothesis, it might well be that he would believe that a limited response to aggression might end hostilities immediately and remove further threat of aggression. Thus the execution of his secondary intention would not be irrational, and the objection to his ability to form it evaporates.

5. AGENT MORALITY AND DETERRENCE

Under a traditional, monistic interpretation of deterrent intention, the problem of agent rationality may well defy resolution. As the toxin puzzle makes clear, one cannot simply intend to perform an act which is irrational. But the dualistic interpretation of deterrent intention, which includes a clearer understanding of the nature and function of I_s, leads to the recognition that the rationality problems of *akrasia* and the toxin puzzle do not translate into rationality problems for deterrence. It also makes clear that the existence of conclusive reasons against acting do not always imply the existence of conclusive reasons against intending to act. At the most we may say that conclusive reasons against action translate into *prima facie* reasons against the

intention which, as is the case with deterrence, may be insufficient grounds on which to convict the agent on a charge of irrationality.

Thus by appealing to the moral justification of the intention to retaliate, we have answered the conceptually prior question of the rational possibility of forming that intention: It is rational (i.e, reason-based) for the deterring agent to form the intention because it is morally justified to do so. He is at once exonerated both morally and rationally for his intention formation.

And finally we may see the point where the supporters and opponents of deterrence must come to a parting of ways. Those who accept the inevitability of escalation must therefore deny the rational possibility of a moral deterrence policy. Those who question escalation (and we saw in Chapter 2 that those ranks are not thin) may look to the arguments presented in this thesis for a defence of the morality of nuclear deterrence. Whether or not the hypothesis is to be accepted is a matter for empirical investigation, the evidence for which we may all be thankful is not, and may never be, available. But within the limited context set out here, deterrence is a morally justifiable endeavour.

The only question remaining then is whether the United States and its NATO allies (or the Soviet Union and its Warsaw pact partners) are justified in engaging in deterrence with the current number of weapons, or whether a reduction in force levels would improve their moral position without jeopardising the peace and security which is the final aim of deterrence. I shall conclude this thesis with a few remarks on that question.

PART VI: CONCLUDING REMARKS

12: TOWARD AN IDEAL DETERRENT

1. THOUGHTS ON AN IDEAL DETERRENT 209

2. CONFIDENCE-BUILDING MEASURES 212

12: TOWARD AN IDEAL DETERRENT

It is with government as it is with medicine, its only business is the choice of evils.

--Jeremy Bentham

Having reached the end of our analysis of intention in nuclear deterrence, we need do no more at this point than remind ourselves of the sometimes circuitous path which that analysis has followed, and of what we have discovered. I shall do that, but I shall also mention something of the moral future of deterrence within the nascent framework of the emerging *perestroika* concerning East-West relations.

I began this thesis with the goal of providing an approach to the analysis of nuclear deterrent intention which might lead to a new moral defence of deterrence. The analysis started with an examination of intention, both as a general concept and as a particular object of critique for a number of commentators on deterrence. As the examination proceeded, it became increasingly clear that the current understanding of deterrent intention was inadequate, failing for example to account for several significant anomalies *vis-à-vis* the ordinary notion of intention. What we needed was a clearer conception of deterrent intention, one which could incorporate and account for the apparent anomalies.

That conception took the form of the dualistic interpretation of deterrent intention. This model was designed to clarify the exact nature of both types of intention inherent in deterrence, and to explain the anomalies which appeared under the monistic view. We then examined the Principle of Double Intention as a possible justification for forming the admittedly immoral intention to retaliate. This analogue to the Principle of Double Effect offered a decision procedure for determining the morality of agents who form immoral intentions, which when applied to nuclear deterrence yielded the conclusion that an agent who endeavoured to deter may be justified in forming the intention to retaliate, provided of course that he could rationally do so.

The examination of that proviso was the subject of Chapter 11, wherein we determined that the intention formation was rationally possible. The deterring agent could indeed form the intention to retaliate, especially if he believed that being forced to act on that intention did not necessarily entail an escalation to all-out nuclear war, a belief which was justified given the warranted doubts about the veracity of the escalation hypothesis. Thus a moral defence of deterrence could be constructed, once the intentions involved had been properly dissected and analysed.

1. THOUGHTS ON AN IDEAL DETERRENT

The main work of the thesis being complete, I shall conclude with a few remarks which may stray into the arena of political science, but which I believe constitute the practical recommendations resulting from the foregoing moral analysis. We begin by noting that there are actually two solutions to the escalation problem which lies at the heart of the rationality question. The first, and the one which we have emphasised, is a theoretical denial of the hypothesis: it might well be the case that

a nuclear exchange will be terminated early, with few destructive detonations. The second solution, and one which we have overlooked until now, is a practical recommendation to eliminate the possibility of nuclear escalation to levels which threaten the kind of global destruction which gives rise to the rationality question. Within the framework of Double Intention, this recommendation constitutes a moral (rather than a political) case for disarmament.

The recommendation is that nuclear powers should seek to disarm to the point where they only retain a minimum effective nuclear deterrent force. This recommendation is motivated by the fact, reflected in our original discussion of deterrence in Chapter 2 (§1.1), that deterrence consists of a credible threat to react to the occurrence of an adversary's unwanted act x by causing him to suffer costs which outweigh the benefits of doing x. If the deterrence policy is effective, the outcome of the adversary's cost-benefit analysis leads him to be deterred from doing x. In order to deter, the policy need only threaten damage sufficient to overcome any benefit to the adversary; it need not threaten more. This is especially true for nuclear deterrence. The comparison with non-nuclear forms of deterrence is well made by Robert Jervis:

> It does not matter which side has more nuclear weapons. In the past, having a larger army than one's neighbor allowed one to conquer it and protect one's own population. Having a larger nuclear stockpile yields no such gains. Deterrence comes from having enough weapons to destroy the other's cities; this capability is an absolute, not a relative one.[1]

Although one may dispute his claim (echoed by Finnis & co.[2]) that deterrence arises from the power to destroy cities, it is certainly true that an 'overkill' capability cannot be strategically justified. Once the sufficient force is attained, no additional weapons

---

[1] Jervis, p. 618. For similar arguments, see also Fisher, p. 89; Hockaday, p. 75; and Kemp, 1987a, p. 279. Against this line of reasoning, it may be argued (see, e.g., Finnis, et al., pp. 211-12) that minimum force level deterrents are destabilising since too few weapons are ineffective and too many are dangerously threatening. However, this argument overlooks the case for mutual force reductions which tend to avoid these problems.

[2] Finnis, et al., esp. pp. 138-39. Their claim that 'city-swapping' is a necessary element of any deterrence policy is one of the cornerstones of their argument against the morality of nuclear deterrence.

are necessary, assuming that deterrence is the sole aim of the agent.

Both the United States and the Soviet Union currently possess many more nuclear warheads than are justified by deterrence. That this fact is generally accepted, not only by critics of deterrence, but by the two superpowers themselves, is demonstrated by recent moves to reach agreement on mutual reductions in strategic and shorter range weapons. These moves to shrink significantly nuclear stockpiles improve the moral case for deterrence in two important ways. First, sharp reductions will eliminate the possibility of a nuclear war escalating to the point where the continued existence of the species is threatened. While this of course would not thereby sanction the use of nuclear weapons, it would mitigate the problem of rationally forming the secondary intention, and thus improve the effectiveness of the deterrent. Secondly, they would remove any doubts about whether those two nations had successfully passed the PDI Countermeasures Test, which was designed to determine if the agent had taken all reasonable steps to ensure against execution of the evil intention. In applying that test to nuclear deterrence, we mentioned that certain classes of weapons run a greater risk of inadvertent or mistaken launch because of their inherent vulnerability to attack. Elimination of those classes would improve the justification of deterrence under the PDI.

While the actual number and types of arsenal reductions are matters for empirical enquiry, and therefore beyond our present scope of discussion, it seems certain that the moral argument we have constructed entails the elimination of some broad types of weapons systems. These would include all immobile land-based missiles, especially those positioned in central Europe and thus highly vulnerable to capture or destruction, as well as the large strategic missiles whose unchanging locations are well known, and thus subject to preemptive attack. Such weapons require their possessor to make hasty (and therefore dangerous) decisions about launch, or risk

their loss. To a lesser extent, those warheads designed for bomber aircraft should also be eliminated. Although they can be recalled, and thus carry with them a built-in delay in execution, they are vulnerable to loss both while on the ground and enroute to their targets.

Of the three basic types of delivery systems, this leaves only submarine-launched ballistic missiles. Because missile submarines are virtually invulnerable to preemptive attack, and will remain so for the foreseeable future[3], they offer both the greatest insurance against inadvertent use, and the lowest risk of escalation, allowing for a reasoned, controlled response to aggression. The only strategic objection to sole reliance on submarine missiles, that their mobile platform results in inaccuracies and thus increases the risk of collateral damage, is rapidly losing its potency as a result of recent technological advances in missile guidance systems.[4]

2. CONFIDENCE-BUILDING MEASURES

The idea of a submarine force as the one power behind deterrence is set forth by Kenny as a 'minimum transitional existential deterrent' leading to eventual complete nuclear disarmament.[5] However, total disarmament does not yet seem to be a realistically attainable goal, given the fact that nuclear weapons cannot be disinvented--the capability will continue to make them accessible--and given the level of mistrust which exists among nations. But there may yet be hope. Although the disinvention problem will remain, trust among nations can be dramatically improved

---

[3] See Hockaday (p. 72), who argues that submarines will remain undetectable barring a major technological breakthrough, which is considered, by British Government assessments, to be remote.

[4] Wohlstetter (1983a, p. 22) goes so far as to argue that the emerging technology 'would permit a conventional weapon to replace nuclear bombs in a wide variety of missions with an essentially equal opportunity of destroying a fixed military target.' See also Fisher (p. 90) who argues for the deterrent sufficiency of submarine missiles.

[5] Kenny, 1985, p.82.

with surprisingly little effort. The opportunities to displace suspicion with trust, and confrontation with cooperation, are abundant. Officially known as confidence-building measures, these chances to improve understanding among opposing nations serve to engender the trust which turns adversaries into partners, and thereby renders the need for deterrence obsolete.

That these measures have a dramatic effect on international relations has been conclusively demonstrated by western reaction to the political changes in the Warsaw Pact nations over the past year. The transformations, first in Poland, and then in East Germany, Czechoslovakia, Romania, and the Baltic and Transcaucasian republics, have been answered with unprecedented offers of aid and assistance by almost all western countries. Examples such as these show that confidence-building measures may well be the key to solving the prisoners' dilemma of the nuclear arms race, a race fueled by mutual distrust.[6]

Throughout this thesis, I have been careful to refer to the primary goal of the deterrence endeavour as 'deterrence of aggression.' But confidence-building measures may unlock the door to a broader goal of 'prevention of aggression' which need not be accomplished by retaliatory deterrence. For, as John Reichart and Steven Sturm remark, 'Deterrence is, after all, not so much an alternative to satisfactory relations as an uncomfortable and perilous burden until their appearance.'[7] Only in an environment of openness and genuine cooperation can the dream of total disarmament reach fruition. In that utopia, a moral justification for deterrence will be moot. Governments which actually seek that day no longer limit themselves to the business of the choice of evils.

---

[6]Gewirth, p. 132.

[7]Reichart and Sturm, p. 152.

BIBLIOGRAPHY

Airaksinen, Timo. 1988. *Ethics of Coercion and Authority.* Pittsburgh: University of Pittsburgh Press.

Alford, Jonathan. 1979. 'Confidence-Building Measures in Europe: The Military Aspects.' *The Future of Arms Control, Part III: Confidence-Building Measures. Adelphi Paper No. 149.* London: International Institute for Strategic Studies. 4-13.

Anscombe, G. E. M. 1966. *Intention.* Ithaca: Cornell University Press.

___. 1968. 'Modern Moral Philosophy.' *Ethics.* Ed. Judith J. Thomson and Gerald Dworkin. New York: Harper & Row. 186-210.

___. 1970. 'War and Murder.' *War and Morality.* Ed. Richard Wasserstrom. Belmont: Wadsworth. 42-53.

___. 1982. 'Action, Intention, and Double Effect.' *Proceedings of the American Catholic Philosophical Association* LVI. 12-25.

Aquinas, St Thomas. 1947. *Summa Theologica.* Trans. Fathers of the English Dominican Province. New York: Benziger Brothers.

Ball, Desmond. 1986. 'The Development of the SIOP, 1960-1983.' *Strategic Nuclear Targeting.* Ed. Desmond Ball and Jeffrey Richelson. Ithaca, New York: Cornell University Press. 57-83.

Begg, David K.H. 1982. *The Rational Expectation Revolution in Macroeconomics.* Oxford: Phillip Allen.

Bennett, Jonathan. 1966. 'Whatever the Consequences.' *Analysis* 26. 83-102.

Bentham, Jeremy. 1970. *Introduction to the Principles of Morals and Legislation.* Ed. J. H. Burns and H. L. A. Hart. London: The Althone Press.

Bok, Sissela. 1978. *Lying: Moral Choice in Public and Private Life.* Hassocks: The Harvester Press.

Boyle, Joseph M. and Sullivan, Thomas D. 1976. 'The Diffusiveness of Intention Principle: A Counter-Example.' *Philosophical Studies* 31. 357-60.

___. 1980. 'Toward Understanding the Principle of Double Effect.' *Ethics* 90. 527-38.

Bratman, Michael E. 1987. *Intention, Plans, and Practical Reason.* Cambridge: Harvard University Press.

Bull, Hedley. 1980. 'Future Conditions of Strategic Deterrence.' *The Future of Strategic Deterrence, Part I. Adelphi Paper No. 160.* London: International Institute for Strategic Studies. 13-23.

Chalfont, The Rt. Hon. The Lord. 1989. 'Arms Control: Opportunities and Dangers.' Oxford University Strategic Studies Group Lecture. 28 November 1989.

Childress, James F. 'Just-War Theories: The Bases, Interrelations, Priorities, and Function of their Criteria.' *Theological Studies* 39. 427-45.

Chisholm, Roderick M. 1970. 'The Structure of Intention.' *The Journal of Philosophy* LXVII:19. 633-47.

Davidson, Donald. 1980. *Essays on Actions and Events.* Oxford: Oxford University Press.

Duff, R. A. 1976. 'Absolute Principles and Double Effect.' *Analysis* 36. 68-80.

Dummett, Michael. 1986. 'The Morality of Deterrence.' *Nuclear Weapons, Deterrence, and Disarmament.* Ed. David Copp. Calgary: University of Calgary Press. 111-28.

Dworkin, Gerald. 1985. 'Nuclear Intentions.' *Ethics* 95. 445-60.

Finnis, John, Joseph M. Boyle, and Germain Grisez. 1987. *Nuclear Deterrence, Morality and Realism.* Oxford: Clarendon.

Fisher, David. 1985. *Morality and the Bomb.* London: Croom Helm.

Ford, John C. 1970. 'The Morality of Obliteration Bombing.' *War and Morality.* Ed. R. Wasserstrom. Belmont: Wadsworth. 15-41.

Foot, Philippa. 1978. 'The Problem of Abortion and the Doctrine of Double Effect.' *Virtues and Vices.* 18-32.

Frey, R. G. 1975. 'Some Aspects to the Doctrine of Double Effect.' *Canadian Journal of Philosophy* V:2. 259-83.

Fried, Charles. 1978. *Right and Wrong.* Cambridge, Mass: Harvard University Press.

Geneva Convention Relative to the Protection of Civilian Persons in Time of War. 12 August 1949.

Gewirth, Alan. 1986. 'Reason and Nuclear Deterrence.' *Nuclear Weapons, Deterrence, and Disarmament.* Ed. David Copp. Calgary: University of Calgary Press. 111-28.

Glover, Jonathan. 1977. *Causing Death and Saving Lives.* New York: Penguin Books.

Grice, Paul, and Judith Baker. 1985. 'Davidson on "Weakness of Will."' *Essays on Davidson Actions and Events.* Ed. Bruce Vermazen and Merrill B. Hintikka. Oxford: Clarendon Press. 27-49.

Grisez, Germain. 1970. 'Towards a Consistent Natural Law Ethics of Killing.' *American Journal of Jurisprudence* XV. 64-96.

___. 1982. 'The Moral Implications of a Nuclear Deterrent.' *The Center Journal* 2. 9-24.

Hardin, Russell. 1986. 'Deterrence and Moral Theory.' *Nuclear Weapons, Deterrence, and Disarmament.* Ed. David Copp. Calgary: University of Calgary Press. 161-93.

Harries, Richard. 1986. *Christianity and War in a Nuclear Age.* Oxford: A. R. Mowbray & Co.

Hart, H. L. A. 1968. *Punishment and Responsibility.* Oxford: Clarendon Press.

___. And Tony Honoré. 1985. *Causation in the Law.* Oxford: Clarendon Press.

Hauerwas, Stanley. 1985. 'Pacificism: Some Philosophical Considerations.' *Faith and Philosophy* II:2. 99-104.

Hobbes, Thomas. 1651. *Leviathan.* Reprinted in *British Moralists.* Ed. D. D. Raphael. Oxford: Oxford University Press. 1969.

Hockaday, Arthur. 1982. 'In Defence of Deterrence.' *Ethics and Nuclear Deterrence.* Ed. Geoffrey Goodwin. London: Croom Helm. 68-93.

Hoffman, Robert. 1984. 'Intention, Double Effect, and Single Result.' *Philosophy and Phenomenological Research* XLIV:3. 389-93.

Honderich, Ted. 1984. *Punishment: The Supposed Justifications.* Harmondsworth: Penguin Books Ltd.

Hughes, Gerard, SJ. 1983. 'The Intention to Deter.' *The Cross and the Bomb: Christian Ethics and the Nuclear Debate.* Ed. Francis Bridger. London: Mowbray. 25-34.

Jervis, Robert. 1979. 'Why Nuclear Superiority Doesn't Matter.' *Political Science Quarterly* 94. Winter 1979/80. 617-33.

Kant, Immanuel. 1956. *Grundlegung zur Metaphysik der Sitten.* Trans. H. J. Paton. New York: Harper & Row.

Kavka, Gregory S. 1978. 'Some Paradoxes of Deterrence.' *The Journal of Philosophy* 75:6. 285-302.

___. 1983. 'The Toxin Puzzle.' *Analysis* 43. 33-36.

Kemp, Kenneth W. 1987a. 'Nuclear Deterrence and the Morality of Intentions.' *The Monist* 70:3. 276-97.

___. 1987b. *Just-War Theory and the Casuistry of Prima Facie Duties.* ms.

Kenny, Anthony. 1966. 'Intention and Purpose.' *The Journal of Philosophy* 63. 642-51.

___. 1973. *Anatomy of the Soul.* Oxford: Basil Blackwell.

___. 1975. *Will, Freedom and Power.* Oxford: Basil Blackwell.

___. 1978. *Freewill and Responsibility.* London: Routledge & Kegan Paul.

___. 1984. 'Better Dead Than Red.' *Objections to Nuclear Defence.* Ed. Nigel Blake and Kay Poole. London: Routledge & Kegan Paul. 12-27.

___. 1985. *The Logic of Deterrence.* Chicago: University of Chicago Press.

Lango, John. 1987. 'Is It Wrong to Intend to Do That Which It Is Wrong to Do?' *The Monist* 70:3. 316-29.

Lee, William T. 1986. 'Soviet Nuclear Targeting Strategy.' *Strategic Nuclear Targeting.* Ed. Desmond Ball and Jeffrey Richelson. Ithaca, New York: Cornell University Press. 84-108.

Mabbott, J.D. 1969. 'Punishment.' *The Philosophy of Punishment.* Ed. H.B. Acton. London: MacMillan & Co. 39-54.

Mackie, J.L. 1967. 'Fallacies.' *The Encyclopedia of Philosophy.* Vol 6. Ed. Paul Edwards. New York: Macmillan Publishing Co. 179-89.

McMahan, Jeff. 1985. 'Deterrence and Deontology.' *Ethics* 95. 517-36.

Mill, John Stuart. 1979. *Utilitarianism.* Indianapolis: Hackett.

Morris, Christopher W. 1985. 'A Contractarian Defense of Nuclear Deterrence.' *Ethics* 95. 479-96.

Murphy, Jeffrie G. 1973. 'The Killing of the Innocent.' *The Monist* 57:4. 527-50.

Nacht, Michael. 1981. 'Toward an American Conception of Regional Security.' *Daedalus* Winter 1981. 1-22.

Novak, Michael. 1982. 'Arms and the Church.' *Commentary* March 1982. 37-41.

___. 1983. *Moral Clarity in the Nuclear Age.* Nashville: Thomas Nelson.

O'Neill, Onora. 1986. 'Who Can Endeavour Peace?' *Nuclear Weapons, Deterrence, and Disarmament.* Ed. David Copp. Calgary: University of Calgary Press. 41-74.

Paskins, Barrie. 1982. 'Deep Cuts Are Morally Imperative.' *Ethics and Nuclear Deterrence.* Ed. G. Goodwin. London: Croom Helm. 94-116.

Peacocke, Christopher. 1985. 'Intention and *Akrasia.' Essays on Davidson Actions and Events.* Ed. Bruce Vermazen and Merrill B. Hintikka. Oxford: Clarendon Press. 51-73.

Pears, D. F. 1985. 'Intention and Belief.' *Essays on Davidson Actions and Events.* Ed. Bruce Vermazen and Merrill B. Hintikka. Oxford: Clarendon Press. 75-88.

Pitcher, George. 1970. '"In Intending" and Side Effects.' *The Journal of Philosophy* LXVII:19. 659-68.

Ramsey, Paul. 1968. *The Just War: Force and Political Responsibility.* New York: Charles Scribner's Sons.

Rawls, John. 1972. *A Theory of Justice.* Oxford: Clarendon Press.

Reichart, John F. and Steven R. Sturm. 1982. *American Defense Policy.* Baltimore: The Johns Hopkins University Press.

Richards, Norvin. 1984. 'Double Effect and Moral Character.' *Mind* XCIII. 381-97.

Sagan, Carl. 1986. 'Nuclear War and Climatic Catastrophe: Some Policy Implications.' *Nuclear War, Nuclear Proliferation and Their Consequences.* Ed. Sadruddin Aga Khan. Oxford: Clarendon Press. 241-75.

Schell, Jonathan. 1982. *The Fate of the Earth.* London: Jonathan Cape Ltd.

Schelling, Thomas C. 1980. *The Strategy of Conflict.* Cambridge, Mass: Harvard University Press.

Shaw, William H. 1984. 'Nuclear Deterrence and Deontology.' *Ethics* 94. 248-60.

Sidey, Hugh. 1985. 'The Presidency.' *Time.* January 28, 1985. 29.

Sidgwick, Henry. 1962. *The Methods of Ethics.* Chicago: University of Chicago Press.

Sterba, James P. 1986. 'Moral Approaches to Nuclear Strategy: A Critical Evaluation.' *Nuclear Weapons, Deterrence, and Disarmament.* Ed. David Copp. Calgary: University of Calgary Press. 75-109.

United States National Conference of Catholic Bishops. 1983. *The Challenge of Peace: God's Promise and Our Response.* Washington, DC: US Catholic Conference.

Walzer, Michael. 1977. *Just and Unjust Wars.* New York: Basic Books.

Warnock, G. J. 1983. *Morality and Language.* Oxford: Basil Blackwell.

Wasserstrom, Richard. 1969. 'On the Morality of War: A Preliminary Inquiry.' *Stanford Law Review* 21:6 June, 1969. 1636-56.

Williams, Bernard. 1984. 'Morality, Scepticism and the Arms Race.' *Objections to Nuclear Defence.* Ed. Nigel Blake and Kay Poole. London: Routledge & Kegan Paul. 99-114.

Wittgenstein, Ludwig. 1958. *Philosophical Investigations.* trans. G. E. M. Anscombe. New York: Macmillan Publishing Company.

Wohlstetter, Albert. 1983a. 'Bishops, Statesmen, and Other Strategists on the Bombing of Innocents.' *Commentary.* June 1983. 15-35.

___. 1983b. 'Letter to the Editor.' *Commentary.* December 1983. 13-22.