1989

Is nuclear deterrence paradoxical deterrence?

Bryan D. Keifer

University of Massachusetts Amherst

Follow this and additional works at: https://scholarworks.umass.edu/theses

Retrieved from https://scholarworks.umass.edu/theses/2484

This thesis is brought to you for free and open access by ScholarWorks@UMass Amherst. It has been accepted for inclusion in Masters Theses 1911 - February 2014 by an authorized administrator of ScholarWorks@UMass Amherst. For more information, please contact scholarworks@library.umass.edu.
IS NUCLEAR DETERRENCE PARADOXICAL DETERRENCE?

A Masters Thesis Presented

by

BRYAN D. KEIFER

Submitted to the Graduate School of the University of Massachusetts in partial fulfillment of the requirements for the degree of

MASTER OF ARTS

September 1989

Department of Philosophy
IS NUCLEAR DETERRENCE PARADOXICAL DETERRENCE?

A Thesis Presented

by

BRYAN D. KEIFER

Approved as to style and content by:

Fred Feldman, Chair

Gareth Matthews, Member

John Robison, Member

John Robison, Department Head
Department of Philosophy
I wish to acknowledge the assistance and encouragement I received from the members of my thesis committee: Fred Feldman, Gareth Matthews, and John Robison. I am particularly indebted to Fred Feldman, who served as the chairman of my committee and read many drafts of this thesis. His comments and ideas were extremely valuable. I also wish to express my appreciation to the Department of the Army for allowing me the opportunity, and the financial support, for graduate study.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGEMENTS ........................................ iii</td>
</tr>
<tr>
<td>LIST OF TABLES ........................................... v</td>
</tr>
<tr>
<td>CHAPTER</td>
</tr>
<tr>
<td>1. INTRODUCTION ........................................... 1</td>
</tr>
<tr>
<td>2. CONSEQUENTIALIST PRINCIPLES ......................... 17</td>
</tr>
<tr>
<td>3. NON-CONSEQUENTIALIST PRINCIPLES .................... 47</td>
</tr>
<tr>
<td>4. CONCLUSION ............................................. 71</td>
</tr>
<tr>
<td>BIBLIOGRAPHY ............................................... 81</td>
</tr>
</tbody>
</table>
LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 2.1: Expected Utility Calculations: Deterrence vs. Disarmament</td>
<td>21</td>
</tr>
<tr>
<td>Table 2.2: Expected Utility Calculations: Retaliation vs. No Retaliation</td>
<td>25</td>
</tr>
<tr>
<td>Table 2.3: Total Expected Utility Calculations: Deterrence vs. Disarmament</td>
<td>38</td>
</tr>
<tr>
<td>Table 2.4: Total Expected Utility Calculations: Retaliation vs. No Retaliation</td>
<td>40</td>
</tr>
</tbody>
</table>
CHAPTER 1
INTRODUCTION

With the use of nuclear weapons against Japan in 1945, the United States initiated a fundamental change in the purpose of military forces. Bernard Brodie, a nuclear strategist of the time, noted that before 1945, military force was intended to fight wars. Once equipped with nuclear weapons, however, the role of military forces could only be to prevent war--most importantly, to prevent nuclear war. Policy makers of the time developed the strategy of nuclear deterrence as a way to achieve this goal. Yet ever since its inception, nuclear deterrence has been controversial. Often debated is the effectiveness of the policy, the credibility of the policy, and the danger of the policy.

Perhaps the most philosophically interesting debate about nuclear deterrence, however, has been over the charge that adoption of a nuclear deterrence policy, together with certain assumptions about the consequences of that policy, results in a moral paradox. This charge is based on the fact that nuclear deterrence rests on the announced intention of the United States (US) to retaliate with nuclear weapons against any aggressor who attacks the US, or its allies, with nuclear weapons. It would seem that this intention to retaliate, if it does deter potential aggressors, and thus prevents nuclear attack, is a morally good intention. However, should deterrence fail, and the circumstances under which retaliation was threatened obtain, then it seems morally wrong to retaliate, since the result would only be a doubling of the nuclear devastation. If it is wrong to retaliate, then it seems that it must be wrong to intend to retaliate.
Thus, the paradox arises from the fact that the intention to do the act seems morally right based on the consequences of forming the intention, but the intention also seems morally wrong, because the intended action is morally wrong. This can also be viewed as a paradox of rationality, if we consider, for example, that on one hand it can be rational to form the intention to retaliate, if the consequences of forming the intention make having the intention rational, but on the other hand, if retaliation is irrational, then it also seems irrational to intend to retaliate. It is these moral and rational paradoxes of nuclear deterrence that I want to examine in this thesis. For convenience, I will refer to deterrence that generates these paradoxes generically as "paradoxical deterrence." When my comments apply to only one of the paradoxes (either moral or rational, but not both), I will so indicate.

To introduce these paradoxes, I first want to trace the development of the nuclear deterrence policy of the US. I then want to explain more fully the circumstances that can make a policy of deterrence "paradoxical deterrence," and then identify certain instances of deterrence under which paradoxical deterrence does not arise. Finally, I will outline how I intend to examine paradoxical deterrence throughout the rest of the thesis.

The debate about the morality of nuclear deterrence has existed almost from the time that the policy was first adopted. Once the incredible power of nuclear weapons was known, it became imperative to deter their use, thus, the policies of nuclear deterrence were born. Over the years, a variety of tensions, hopes, fears, perceptions, and technological developments have determined the shape of these policies.
At the end of WWII, the United States possessed a nuclear monopoly and proceeded to demobilize much of its conventional armed forces, while the Soviets continued to maintain a very large military. The first seeds of tension were planted during this period as the US continued to develop its nuclear arsenal in order to ease its fears of the large Soviet military force, while at the same time, the Soviets continued to maintain their large military in an attempt to counter the growing US nuclear superiority. By the mid-1950s, however, Soviet nuclear development had progressed to the point that they could credibly threaten the US homeland. The US no longer possessed a nuclear monopoly, although it did possess a technological advantage.

After the Korean war, the Eisenhower administration developed a strategy designed to rely on nuclear weapons in lieu of a large conventional (non-nuclear) force. This policy became known as "massive retaliation." It threatened that, in the event of Soviet aggression, the US would devastate the Soviet homeland with nuclear weapons.

Two criticisms of that policy are still pertinent for the philosophical debate today. First, it was pointed out that the Korean war, a non-nuclear war, had taken place during a period of US nuclear monopoly. Thus, it seemed that there were very real limits to the types of aggression that nuclear weapons could deter. Second, concerns were raised that the massive, indiscriminate counterattack against the Soviet homeland threatened under "massive retaliation," would be immoral and thus, not believable or credible. If the Soviets did not believe that the US would actually "massively retaliate," then the policy would have no deterrent value. This criticism was important for two reasons: First, it raised the important question of the morality of
nuclear retaliation. Second, it identified the fact that for nuclear deterrence (and perhaps any type of deterrence) to be effective, it must be credible.

Although the question of whether retaliation using nuclear weapons was immoral remained, strategists did attempt to increase the credibility of the nuclear threat by developing strategies that involved the concepts of "limited nuclear war" and "counterforce targeting."

The concept of "limited nuclear war" stressed the ability to use nuclear weapons in selective ways against specific targets rather than relying on the all-out strategic retaliation threatened under "massive retaliation." In theory, the threat to use smaller, faster, and more accurate nuclear weapons, that could be targeted against primarily military targets, would be more credible than a threat of massive retaliation against population centers. It was thought that if the threat to retaliate could be made more credible by this new targeting policy, then it would have greater deterrent value.

However, difficulties were immediately apparent with this strategy. First, it introduced another of the paradoxes of deterrence theory: that nuclear war was to be made less likely by making it more likely. Specifically, the difficulty with this line of reasoning was that one could not be sure that it was the credibility of the deterrent that would increase, rather than the actual likelihood of nuclear war. It wasn't clear that a limited nuclear war could be kept limited. It seemed reasonable to expect that if one side began to perceive that it was losing a limited nuclear exchange, then it would begin to employ larger weapons in a less discriminating way. It seemed that the net result of the limited war strategy was to remove some inhibitions to the conduct of nuclear war without providing any way to keep nuclear war within any kind of limits.
These same concerns about the credibility of "massive retaliation" also led to the formulation of the distinction between "countervalue" targeting and "counterforce" targeting. "Countervalue" targeting is targeting of nuclear weapons against cities and industries regardless of their military importance. "Counterforce" targeting, on the other hand, is the targeting of purely military targets, in particular, the nuclear weapons of the other side. The massive retaliation policies of the 1950s were countervalue.

During the 1960s, counterforce policies were considered in order to relieve a number of tensions that arose from policies of countervalue. Credibility was certainly a key consideration. As discussed above, retaliation against military targets posed a much more credible threat to the Soviets than did massive retaliation (countervalue). Counterforce was also intended to serve the purpose of damage limitation to the US homeland. It was thought that if the US announced war plans that called for retaliation against Soviet military targets rather than Soviet cities, the Soviets would respond in kind and not attack US cities.

By the mid-1960s, however, it was clear that counterforce policies would not solve these problems. First, it was uncertain that the Soviets would follow a policy of not attacking US cities if the US did not target Soviet cities. Second, it became apparent that counterforce policies, despite possible gains in credibility, were perhaps more unstable than countervalue policies. This followed from the fact that if the Soviets knew that their military and missiles were the primary targets of the US, then they would have a greater fear of a US surprise attack aimed at destroying their missiles. To preclude this possibility, the Soviets might be tempted to initiate a preemptive attack.
in order to protect their own missiles. The US would then fear such a preemptive strike also, and the situation would become very unstable.

This instability can also drive a costly arms race as each side seeks advances in technology that will increase the speed and accuracy of their weapons in order to prevent their destruction under a surprise attack. Each side may also be driven to installing automatic systems that launch missiles as soon as launch of the opponent's missiles is detected. Such systems tend to put the launch of missiles on a hair trigger, increasing the chances of accidental or mistaken missile launches, which will further increase tensions and destabilize the strategic balance. As a result of these concerns, the US, by the mid-1960s, had returned to policies of countervalue deterrence, which had then became known as "mutually assured destruction" (MAD).

The US strategic doctrine has remained essentially the same from the adoption of MAD through the 1980s. One significant difference, however, is that there is now a mix of countervalue and counterforce weapons, and weapons that can serve either role. Whether these weapons will serve to create instability between the nuclear powers remains to be seen.

Since the US still maintains a strategy based primarily on MAD, many of the same problems still exist: Is the threat of retaliation against the Soviet population morally permissible? Is such a threat credible? Does such a threat actually deter aggression? If so, what types of aggression, i.e., conventional aggression and nuclear aggression, only nuclear aggression, nuclear blackmail, etc?

I think that it's been clear over the years that our policies of nuclear deterrence do not serve to deter conventional aggression. The US
involvement in Korea occurred while the US had a virtual monopoly on nuclear weapons. Conventional aggression by North Viet Nam was certainly not deterred by the US nuclear arsenal. Continued Soviet occupation of eastern Europe after WWII, and military actions in Hungary, Czechoslovakia, and Afghanistan were also undeterred by US nuclear weapons and policies of nuclear deterrence. It seems that if nuclear deterrence deters anything, it can only be nuclear aggression or nuclear blackmail (i.e., threatening use of nuclear weapons in order to force another government to submit to some political or economic demands). Some strategists argue that nuclear deterrence does not even deter these behaviors and that the lack of nuclear aggression and/or blackmail during the nuclear age can be shown to be the result of other factors that have influenced strategic decisions.

Determining exactly what it is that nuclear deterrence does deter, if it deters anything, is a very difficult question. Although it's a difficult determination to make, the answer may be very important to the normative evaluation of nuclear deterrence. For example, it would be important to know what types of aggression nuclear deterrence will deter if one is making a normative evaluation of nuclear deterrence that appeals to the consequences of adopting the deterrent policy. The more types of aggression the policy deters, then the better the expected consequences of the policy. I think that it's very unclear whether a policy of nuclear deterrence can deter conventional aggression. Therefore, for my purposes, I intend to assume that nuclear deterrence has no deterrence value against conventional aggression. Thus, when I discuss the benefits and hazards of nuclear deterrence, I am assuming that nuclear deterrence policies only deter nuclear aggression and/or nuclear blackmail.
At the outset, I said that I wanted to concentrate on "paradoxical deterrence." I explained that paradoxical deterrence arises from the cases where it seems morally right to have the intention to retaliate (based on the consequences of forming the intention), but also seems morally wrong to form the intention to retaliate because retaliation itself is morally wrong. It should be noted that not just any case of deterrence is "paradoxical deterrence." Thus, we need to specify more explicitly what constitutes "paradoxical deterrence," or at least identify those cases of deterrence where the paradox does not arise.

One of the important features that seems required for paradoxical deterrence to arise is that the intention to retaliate is a *sincere* intention, and that the defender perceives that having the intention is the only way to prevent a terrible disaster (a nuclear first strike), even though to carry out the intention is to do something morally wrong. Gregory Kavka's *Special Deterrent Situations* (SDS), which correspond to my "paradoxical deterrence", note this requirement. He says that an agent in an SDS believes that the following conditions hold:

"First, it is likely he must intend (conditionally) to apply a harmful sanction to innocent people, if an extremely harmful and unjust offense is to be prevented. Second, such an intention would very likely deter the offense. Third, the amounts of harm involved in the offense and the threatened sanction are very large and of roughly similar quantity (or the latter amount is smaller than the former). Finally, he would have conclusive moral reasons not to apply the sanction if the offense were to occur." ¹

Kavka's first and second conditions appeal to the characteristics that the intention must be a real intention and that the agent must believe that having the intention will actually deter the offense, in order for "paradoxical deterrence" to arise. Kavka notes that deterrence depends importantly on the potential wrongdoer's beliefs about the prospects of the sanction being applied. Thus, if the agent could successfully influence the potential wrongdoer's beliefs by bluffing, i.e., by publicly announcing the intention to retaliate while secretly intending not to retaliate if the offense is committed, and if the potential wrongdoer did not detect the bluff, then we would say that the agent had successfully deterred the offense.

Such a strategy as bluffing would be characterized as deterrence (since the offense may be successfully deterred), but not as "paradoxical deterrence." For in the case of bluffing, the agent actually has the intention not to retaliate, although he has publicly announced that he will retaliate. Thus, no paradox arises since the agent is not sincerely intending to carry out a morally wrong action.

It may be that bluffing would be successful in deterring the potential wrongdoer. In fact, it may be that the best action for the US to pursue would be some type of bluff that did not entail forming the sincere intention to retaliate. This helps to illustrate another point, hinted by Kavka's second condition above: The agent must believe that the best way to deter the potential wrongdoer, or the way with the greatest chance of success, is to form the sincere intention to retaliate.

It seems that "paradoxical deterrence" requires that the situation be such that the best (morally right, rational, util. maxing, etc.) course of action requires the agent to form the sincere intention to retaliate. If it was the case that the agent determined that bluffing would be the best method of
deterrence, then he should bluff. If the best course of action requires bluffing, then it could not be morally right or rational to have the sincere intention to retaliate. Thus, the paradox would not arise.

If we consider for a moment whether a policy of bluffing would be better than the current US policies, we should note that many nuclear strategists have claimed that the US would never be able to maintain successfully a deterrence policy that relied on a bluff because we could never keep secret the fact that we didn't really intend to retaliate. Herman Kahn has written that our deterrent policy would not be credible,

"...unless we really intend to do it. If we are only pretending that we would do it, the credibility and therefore the deterrent value of our force is almost certain to be lessened by the automatic and inevitable leaks. While we can probably keep the details of our war plans secret, it is most unlikely that we can keep the philosophy behind them secret." 2

Obviously, if the Soviets learned that our deterrent policies rested on a bluff, and that our plans did not call for a retaliatory strike, it is likely that our policies would have little deterrent value. Since the procurement and modernizing of our nuclear arsenal is driven largely by our strategy, and since this procurement and modernizing requires the work of a vast number of contractors and subcontractors, I suspect that Kahn is correct in asserting that leaks would be inevitable.

Nevertheless, there might be any number of other methods that would be better than our current policies of nuclear deterrence or there may be other reasons that independently make forming the intention wrong. David Lewis has noted that,

"It might be wrong for independent reasons to form the deterrent intention. For it might be too risky; it might be unlikely to succeed; it might carry other costs, e.g., in damaging the relationship between the parties. Or there might be a better means of dissuasion available. It might be possible to deter without forming a conditional intention to retaliate: by pretending to have the intention, by making retaliation automatic... or simply by leaving it uncertain what might happen."3

For paradoxical deterrence to arise, it must be the case that forming the intention to retaliate is determined to be a morally right action. If there are other possible courses of action that are better, or if forming the intention is wrong for independent reasons, then the paradox does not arise.

Likewise, paradoxical deterrence requires that the act which is the object of the intention, the act of retaliation, must be determined to be morally wrong. From this, it is claimed that the intention to retaliate is wrong because retaliation itself is wrong. Thus, if retaliation were morally right, then the strategy might still correctly be called deterrence, but it would not constitute "paradoxical deterrence." For in such a case, the intention, and the doing of the action, would both be morally right, and thus, no paradox arises.

How might this occur? It might be the case that US retaliation after a Soviet strike does not necessarily require a total, indiscriminate nuclear strike against the the Soviet homeland. Perhaps the US response would serve to destroy only the remaining Soviet missiles in order to prevent a second strike by the Soviets. A US retaliation might also serve to prevent the Soviets from future world domination. It might also be that a retaliatory strike by the US would serve to end a continuing nuclear conflict. If any of these

possibilities were true, and retaliation were the best course of action, then the paradox would not arise.

It also seems reasonable to exclude from "paradoxical deterrence" those cases of deterrence where the morality of the intention or the retaliation is dependent on external factors unrelated to the deterrence of nuclear aggression. David Gauthier\(^4\) has suggested such an exclusion in his discussion about the credibility of the intention in "paradoxical deterrence." Gauthier is concerned that the credibility of the agent's threat to retaliate is lessened if the potential wrongdoer knows that the agent believes the act of retaliation is not rational or is morally wrong. However, he thinks we should be careful to insure that we reject cases where other, external circumstances could be arranged so as to increase the credibility of the threat. Consider the following examples:

Suppose the President of the US adopts a nuclear deterrent policy, but believes that retaliation is morally wrong. There might be situations where an external factor comes into play, perhaps that the President also believes that he ought to be a man of his word. In fact, it might be that the President feels so strongly about being a man of his word, that even though honoring his expressed intention would lead to retaliation, he would prefer to retaliate rather than not to honor his expressed commitment.

Another example might be a case where the President has made a side bet that he would lose if he failed to retaliate under the specified conditions.

His gain from winning the bet might be sufficient to make retaliation his best course of action, in spite of the negative consequences of retaliation.

I think that these cases of promise keeping and side bets are examples of types of cases that should be rejected as instances of "paradoxical deterrence." In these types of cases, the situation is designed so that the morality and rationality of the intention, or of the retaliation, is determined by some factor that is completely unrelated to the nuclear deterrence problem. I'm not interested in those cases, for example, where the benefits from winning a side bet are what make retaliation the best course of action. What I want to consider are the cases where the morality and rationality of the intention and the retaliation are determined by issues commonly associated with nuclear deterrence.

Gauthier also suggests that there might be cases where the President's expressed intention is actually a warning that retaliation may be beyond his control. For example, suppose the President were an imperfectly rational individual. During his thoughtful, reflective periods he may realize that retaliation is an irrational course of action. Yet if the Soviets were to attack, his considered preference for not retaliating may be overcome by anger, rage, or panic in such a way that he might lose control and retaliate.

Likewise, suppose that in expressing the intention to retaliate the President was actually warning the Soviets that a complex, irreversible, automatic retaliation machine had been activated which would automatically retaliate if the Soviets attacked. In this case, it seems that the President is simply warning the Soviets that he no longer has control over the retaliation; it will occur automatically if they were to attack.

The two examples above also illustrate another type of case that I do not consider to be "paradoxical deterrence." In these cases, the situation has
been constructed so as to take the act of retaliating out of the hands of the agent or to make the retaliation an act of an irrational agent. I want to consider those cases where the agent is fully rational, where the agent forms the genuine intention to retaliate, and where it is fully in the power of the agent to do or not do the retaliation. The cases where the agent may retaliate during a wave of irrationality, or where a machine is activated over which he has no control, may be cases of deterrence, but I do not consider them to be instances of the "paradoxical deterrence."

I think that the above discussion illustrates four limiting conditions that a case of deterrence must meet in order to be "paradoxical deterrence." First, it must be the case that forming the intention is the best course of action. If forming the intention is wrong for independent reasons, then the paradox does not arise. Second, it must be the case that carrying out the intention, and actually retaliating, is morally wrong. If there are independent factors that make retaliation morally right, then the paradox does not arise. Third, "paradoxical deterrence" does not include those cases where the morality of the intention or the retaliation depends on factors that are completely unrelated to the deterrence situation (e.g., side bets or beliefs in promise keeping). Fourth, "paradoxical deterrence" does not include those cases where the expressed intention is actually a warning to the potential offender that the act of retaliation is beyond the control of the agent.

With these conditions in mind, we can move on to considering whether the nuclear deterrence policy of the US is a case of "paradoxical deterrence", and if so, whether the paradox can be solved. I have defined "paradoxical deterrence" as the instances of nuclear deterrence where there are two lines of reasoning about the normative status to retaliate: First, is the line of
reasoning which shows that forming the intention to retaliate is morally right, and rational, based on the consequences of forming the intention (primarily the consequence of preventing nuclear war). Second, is the line of reasoning which shows that forming the intention to retaliate is wrong because retaliation itself is wrong, or irrational. This line of reasoning uses a moral principle which holds that it is wrong to intend to do something that is wrong.

In Chapter 2, I want to examine more closely the first line of reasoning that is based on an assessment of the consequences of forming the intention to retaliate. I will present and explain some commonly discussed principles of morality and rationality, such as the Maximization of Expected Utility, Disaster Avoidance Principle, Maximization of Total Expected Utility and Maximin, that utilize consequences in assessing the normative status of an action. I will suggest applications of these principles to the case of nuclear deterrence, to see if they will plausibly support the first line of reasoning.

In Chapter 3, I will examine the second line of reasoning which holds that the intention to retaliate is wrong because retaliation itself is wrong. This line of reasoning depends on non-consequentialist principles that determine the normative status of an action by appeal to features of the action other than its consequences. This line of reasoning appeals to a non-consequentialist argument to show that nuclear retaliation is wrong, and then uses the Wrongful Intentions Principle (WIP), which holds that it is wrong to intend to do an act that is itself wrong, to show that the intention to retaliate is wrong. In this chapter, I will examine the non-consequentialist principles used to show that nuclear retaliation is wrong, and then focus
specifically on the WIP, to determine its usefulness and applicability to the case of nuclear deterrence.

Finally, in Chapter 4, I want to draw some conclusions about "paradoxical deterrence" and the nuclear deterrence policy of the US.
Paradoxical deterrence seems to be rooted in the following argument:

1. Retaliation is irrational.
2. If retaliation is irrational, then forming the conditional intention to retaliate is irrational.
3. Therefore, forming the conditional intention to retaliate is irrational.

The argument appears to be sound. Yet many people believe that under present circumstances, adoption of a nuclear deterrence policy, which includes forming the conditional intention to retaliate, is the best way to prevent a nuclear war, and thus is a completely rational course of action. A conditional intention to do an act, A, is the intention to do A if a certain specified condition were to occur. In the case of nuclear deterrence, the act is US nuclear retaliation and the specified condition is a Soviet nuclear first strike against the US. Thus, for the US to have the conditional intention to retaliate means that it intends, if the specified condition occurs (the first strike), to do the act (nuclear retaliation).

We are left in the seemingly paradoxical position of asserting that forming a conditional intention to retaliate is rational based on its consequences, while also accepting the argument above that concludes that forming the conditional intention to retaliate is irrational because retaliation is irrational.

The paradox seems to arise in the same way when we consider morality. On one hand, the conditional intention to retaliate seems morally right if it
prevents war, and on the other, the conditional intention to retaliate seems morally wrong if retaliation itself is morally wrong.

The paradox arises from application of these different lines of reasoning to evaluate the formation of the conditional intention. The argument above concludes that the conditional intention is irrational based solely on the irrationality of retaliation itself, disregarding possibly desirable consequences from forming the conditional intention (I will call this the non-consequentialist line of reasoning). The second line of reasoning evaluates the conditional intention on the basis of the consequences of forming the intention, without regard to whether the action intended is rational or irrational (I will call this the consequentialist line of reasoning). The paradox forces us to examine more closely the application of each of these lines of reasoning to see if they are being applied consistently and appropriately.

In this chapter, I want to examine the consequentialist line of reasoning that leads to the conclusion that formation of the conditional intention is rational or morally right. I will show that when we apply some of the most widely discussed principles of rationality, and principles of morality, to the case of nuclear deterrence, we will find that it is consistent with these principles for the conditional intention to retaliate to be rational (moral), even though retaliation itself is irrational (immoral) according to the same principle.

I will first consider the rationality of forming the conditional intention to retaliate. To illustrate this, I want to suggest applications of the principles of Maximization of Expected Utility, Disaster Avoidance, and Maximin to the case of nuclear deterrence. I think we will find that it is consistent with each
of these principles for the formation of the conditional intention to retaliate
to be rational while the act of retaliating is irrational. These results suggest
that, under certain circumstances, it may be rational to intend to be
irrational.

Maximization of Expected Utility

I first want to consider the principle of Maximization of Expected Utility
(MEU), which can be formulated as follows:

**MEU**: An action is rational for a subject, S, if and only if it maximizes S's expected utility.

In this discussion, I am concerned with the decisions and preferences of
the US. Therefore, according the MEU, an action would be rational for the US
if it maximized the utility of the US.

Application of MEU to the case of nuclear deterrence requires that one
determine the possible outcomes from each alternative action open to the
United States. We then determine the conditional probability (between 0-1)
for each of these outcomes, such that the sum of the probabilities of all the
outcomes of an action equals 1. We then assign a utility value to each of the
outcomes. For each outcome of an action, we multiply its utility by its
probability on the action. The sum of these products of the probabilities
times the values for all of the outcomes of an action is the expected utility for
that action. The rational action according to MEU is that action such that no
alternative action has greater expected utility. I will limit my discussion of
alternative actions open to the US to only two: forming the conditional
intention to retaliate (which requires stockpiling nuclear weapons) and nuclear disarmament.

I believe that the relevant outcomes for the action of adopting a nuclear deterrent policy are the following: 1. Soviet first strike; 2. Soviet nuclear blackmail; 3. Nuclear Peace; 4. Accidental US first strike.

I suggest "Soviet first strike" instead of "nuclear war" for the following reason. The issue is whether or not to adopt a deterrent policy which threatens the Soviets that if they use nuclear weapons against the US, then the US will respond with nuclear weapons against the Soviet Union. Kavka observes that such deterrent policies consist of "those conditional intentions whose existence is based on the agent's desire to thereby deter others from actualizing the antecedent condition of the intention."\(^5\) US deterrence policy is intended to deter the Soviets from "actualizing the antecedent condition of the intention," i.e., it intends to deter them from launching a nuclear first strike. While a nuclear deterrence policy specifically threatens retaliation, neither the intended purpose of the policy, nor the success or failure of the policy requires fulfilling the intention and retaliating. The policy is simply intended to deter the Soviets from creating a set of circumstances such that the US is forced to decide whether to launch its missiles at the Soviets.

Short of an actual first strike, it would be possible for the Soviets to threaten use of their nuclear weapons to dictate economic or foreign policy terms that are not in the best interest of US. Thus, the possibility of nuclear blackmail is a relevant outcome. Finally, we must note that adoption of a deterrent policy assumes the stockpiling of nuclear weapons readily

\(^5\)Kavka, 290

20
available for retaliation. Such stockpiling of weapons by the US also brings with it the chance of a US accidental launch, which could bring about a Soviet counter-attack. This potential for a Soviet counter-strike is of negative value to the US and should be considered as an undesirable outcome of the policy. The chance of a Soviet accidental launch would be included in the chance of a Soviet first strike.

The outcomes for the policy of nuclear disarmament would be the same as that of deterrence, except that US accidental launch would not be included (the US would no longer maintain nuclear weapons under a policy of nuclear disarmament).

There is great debate about the probabilities and values that can be assigned to these outcomes. However, I think that the following assignment in Table 2.1 would be plausible:

Table 2.1
Expected Utility Calculations: Deterrence vs. Disarmament

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Probability</th>
<th>Value</th>
<th>PxV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Soviet First Strike</td>
<td>.1</td>
<td>-1000</td>
<td>-10</td>
</tr>
<tr>
<td>US Accident</td>
<td>.05</td>
<td>-500</td>
<td>-25</td>
</tr>
<tr>
<td>Nuclear Peace</td>
<td>.75</td>
<td>1000</td>
<td>750</td>
</tr>
<tr>
<td>Blackmail</td>
<td>.1</td>
<td>-750</td>
<td>-75</td>
</tr>
<tr>
<td>Deterrence</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blackmail</td>
<td>.4</td>
<td>-750</td>
<td>-300</td>
</tr>
<tr>
<td>Nuc. Disarm. Blackmail</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nuclear Peace</td>
<td>.4</td>
<td>1000</td>
<td>400</td>
</tr>
<tr>
<td>Total:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total: +550</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Soviet First Strike</td>
<td>.2</td>
<td>-1000</td>
<td>-200</td>
</tr>
<tr>
<td>Nuc. Disarm. Blackmail</td>
<td>.4</td>
<td>-750</td>
<td>-300</td>
</tr>
<tr>
<td>Nuclear Peace</td>
<td>.4</td>
<td>1000</td>
<td>400</td>
</tr>
<tr>
<td>Total: -100</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Estimating probabilities of Soviet choices of actions is a very difficult exercise. Yet we do know that the Soviets tend to be very cautious with their
military power, particularly with nuclear weapons, and very protective of their homeland. Thus, I think that the chance that the Soviets will launch a nuclear first strike in spite of US threats of retaliation, and thereby deliberately invite a nuclear counter-strike against their nation, is very small. Likewise, it seems to me that a nuclear blackmail attempt by the Soviets against a United States that also possessed nuclear weapons would have very little chance of success. Thus, I suggest that the probability of Soviet nuclear blackmail under deterrence is very small.

I think that the chance of an accident or mistake that resulted in a US first strike is also very small. The safety systems built into nuclear weapons, and the fact that multiple persons must act in unison to launch missiles reduces greatly the chance of accidental launch.

I also think that the chance that the US might mistakenly believe that it is under attack and launch its missiles in order to prevent them from being destroyed on the ground is very small. The US does not have to fear that all its missiles could be destroyed by a Soviet surprise attack due to the large number that are deployed on submarines. The relative invulnerability of submarines means that the President has some flexibility in waiting for confirmation of warnings of Soviet attacks. Therefore, it seems to me that there would be only a very small chance of missiles being launched in response to mistaken detections of Soviet missile launches.

Soviet first strike, nuclear blackmail, and an accidental US launch account for 25% of outcome probability on the action. Thus, the remaining outcome, nuclear peace, can be set at .75.

If the threat of retaliation keeps the chance of Soviet first strike low, then removal of that threat may increase the chance of a Soviet first strike.
Thus, I have rated the chance of a Soviet first strike under disarmament as higher than under deterrence. I also suggest that the possibility of Soviet nuclear blackmail, under US disarmament, is somewhat higher than that of a first strike. In some circumstances, nuclear blackmail may be a more effective way for the Soviets to pursue their goals than launching nuclear missiles at the US. And in most cases, a US without nuclear weapons would be compelled to submit to the Soviet demands rather than risk even a limited nuclear attack.

In assigning values, I'm suggesting that failure of the deterrent policy is as bad as success of the policy is good. Nuclear blackmail by the Soviets, which might include restriction of US free markets, the restriction of free action by the US government, and threatened destruction of US military facilities is rated very bad, but not as bad as a nuclear strike. Loss of US freedoms is considered by many to be a catastrophic outcome, yet I don't see it being as bad for the US as the nuclear destruction and long term nuclear effects that would result from being the victim of a nuclear attack.

I assume that an accidental nuclear strike by the US would create undesirable consequences for the US that would be half as bad as a first strike by the Soviets. An accidental US strike would be much smaller than a planned attack and we might expect the Soviet response to be proportional to the attack. I'm also assuming that communication between the US and Soviets would allow for warning and explanations of an accidental or mistaken launch that would serve to reduce the Soviet response.

The value of these outcomes is held the same for the policy of nuclear disarmament.

The calculations from Table 2.1 show that under the circumstances I have suggested, adoption of a deterrence policy results in greater expected
utility for the US than does a policy of nuclear disarmament. Thus according to MEU, adoption of the deterrent policy is the rational course of action.

There may be debate about these calculations, but to make my point, I only need to assume that these figures are reasonably plausible and could be true.

Now suppose the Soviets attack in spite of our threat of retaliation. We are now faced with a new decision problem that includes possible actions that were not part of the earlier calculations. The fact that the Soviets have attacked presents two new possible actions to be considered: retaliation and no retaliation. Having chosen deterrence previously, we can assume that the US possesses the capability to launch a retaliatory strike. Thus, this is an available action. The relevant choices now open to the US are retaliation and no retaliation.

It seems to me that the relevant outcomes to the action of retaliation would be the following: Retaliation might result in an unacceptable increase in the nuclear devastation. By retaliating, the US would be doubling the total amount of world-wide nuclear damage. Additionally, retaliation by the US may invite a second nuclear strike by the Soviets, further increasing the damage to the US without effectively ending the conflict. This outcome would also include the possibility that once the Soviets had launched multiple nuclear attacks against the US, it would attempt domination of what remained of the US. On the other hand, retaliation might serve to end the war quickly. A strike against the Soviets might destroy their remaining missiles or break their resolve to fight. The Soviets might also stop fighting in order to deal with their own nuclear devastation. Thus, retaliation might serve to end the war in essentially a stalemate between the US and the Soviets, rendering each unable to dominate the other.
One outcome relevant to non-retaliation also seems to be that it might end the war. If the US did not strike back, the Soviets may simply declare victory and cease their aggression. This outcome assumes reduced nuclear devastation because the nuclear exchange would be confined to the first strike by the Soviets. Another outcome possible if the US did not retaliate is that the one-sided Soviet nuclear attack would be followed by attempted Soviet domination, primarily through what would now be very credible threats to the US of nuclear strikes if Soviet demands were not met. Assuming these are the significant outcomes, we might assign probabilities and values as shown in Table 2.2:

Table 2.2
Expected Utility Calculations: Retaliation vs. No Retaliation

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Probability</th>
<th>Value</th>
<th>PxV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retaliation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>More destruction</td>
<td>.55</td>
<td>-1000</td>
<td>-550</td>
</tr>
<tr>
<td>End the war</td>
<td>.45</td>
<td>+1000</td>
<td>+450</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>+100</td>
</tr>
<tr>
<td>Total:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No Retal.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>End the war</td>
<td>.6</td>
<td>+1000</td>
<td>+600</td>
</tr>
<tr>
<td>Soviet domination</td>
<td>.4</td>
<td>-500</td>
<td>-200</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>+400</td>
</tr>
<tr>
<td>Total:</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In the outcomes for retaliation, I'm suggesting the probability that retaliation will result in increased nuclear destruction and the possibility of Soviet domination is greater than the probability that retaliation would serve to immediately end the war (e.g., by destroying all the remaining Soviet missiles). If US retaliation did end the war quickly, however, both the US and the Soviet Union would probably have enough damage that there would be little chance that either would attempt to dominate the other or the world.
If the US did not retaliate, the chances of ending the war and the chances of Soviet domination may be fairly close. If the US failed to retaliate, then it seems that there would be little purpose in the Soviets launching more nuclear strikes. On the other hand, if the US did not respond to a Soviet first strike by retaliating, we might expect the Soviets to be much more likely to attempt domination of US policy or to use nuclear blackmail against the US to further their goals because they would no longer fear US retaliation.

I'm assuming that the negative value to the US of Soviet domination after retaliation (included in the outcome of "more destruction"), and the outcome from no retaliation, is less than it was in our first calculation. Once portions of the US have suffered a nuclear strike, threats of domination by the Soviets will seem less significant than if they were directed against an undamaged US. We can see from the calculations above, that if the Soviets did launch a first strike, the rational action for the US, according to MEU, would be no retaliation.

The first MEU calculation (Table 2.1) was made to determine the action which would be most successful in keeping the Soviets from launching a nuclear first strike. If the Soviets launched a first strike in spite of our policy of deterrence, then we face different circumstances and a different decision about how to respond.

Thus, the second set of calculations in Table 2.2 reflects the new decision problem faced by the US, given the occurrence of the Soviet first strike. The second set of calculations is about new possible actions (retaliation and no retaliation) that were not considered in the first set of calculations. My
description of circumstances in the second set of calculations was intended to show that it is consistent with MEU for there to be circumstances such that formation of the conditional intention to retaliate is rational, but should the specified conditions obtain, then MEU will show that retaliation itself is irrational. Thus, there can be very plausible combinations of circumstances such that it is consistent with MEU for it to be rational to intend an action that would be irrational to perform.

Disaster Avoidance Principle

Another commonly suggested principle for determining rationality is the disaster avoidance principle. Gregory Kavka has suggested a formulation of the DAP that will be useful to my discussion here. His formulation of the Disaster Avoidance Principle (DAP) is the following:

DAP: when choosing between potential disasters under two-dimensional uncertainty, it is rational to select the alternative that minimizes the probability of disaster occurrence.6

In his application of DAP to nuclear deterrence, Kavka considers the same two actions that were considered in the discussion of MEU: adoption of a policy of deterrence and unilateral nuclear disarmament. He argues that each of these actions may lead to a disaster. The practice of deterrence against the Soviets may be unsuccessful and result in a Soviet first strike.7

6Kavka, Gregory, “Deterrence, Utility, and Rational Choice, Theory and Decision, 12, 1980, p. 43
7Kavka describes the potential disaster from deterrence as nuclear war, rather than specifically a Soviet first strike, implying that the US must retaliate as part of a policy of deterrence. I believe that the US would have a choice about retaliation, even after adopting a deterrent policy, thus in my discussion I will consider the potential disaster from
Unilateral nuclear disarmament by the US may lead to world domination by the Soviets. Thus, we are forced to decide between these two potential disasters.

When Kavka says that the decision is made under two dimensional uncertainty, he means to indicate the fact that when the decision is to be made, the agent does not have either (a) reliable estimates of the probabilities of outcomes of the actions or (b) reliable utility estimates for the outcomes of the actions. Kavka argues that,

"Assigning numerical probabilities to the possible outcomes of the available choices would require selecting an appropriate reference class of past situations to provide data on the relative frequencies of various outcomes."\(^8\)

There is no unique reference class of past situations that can provide relative frequencies of a Soviet first strike, or the frequency of Soviet domination given US unilateral nuclear disarmament, that would provide the evidence we would need in order to make accurate estimates of the probabilities of these outcomes. Nor is there a scientific theory of international relations that could predict these probabilities in some indirect way. It seems unlikely that we could get reasonable people to agree on any set of probabilities for the outcomes of a Soviet first strike and Soviet domination.

Likewise, there is great difficulty in assigning utility values to the outcomes. There is no empirical data on the effects of large scale nuclear attacks or on world domination by an opposing power, so it is extremely deterrence to be a Soviet first strike. I think that Kavka's arguments reach the same conclusion regardless of which of these disasters is specified in the application of DAP.\(^8\) Kavka, p. 43
difficult to estimate with confidence the utilities of the outcomes. The task of defining a common measure of utility that can be meaningfully assigned to such varied outcomes as nuclear war, loss of freedom, nuclear blackmail, etc. is itself particularly difficult.

Although we can't estimate the utilities and probabilities with confidence, Kavka argues that we can determine an ordinal ranking of the values of the outcomes and the probabilities of the outcomes. He claims that we can judge that the death and destruction from a nuclear attack would be a worse disaster for the United States than would domination by the Soviets.

Kavka also claims that we can judge that the chance of nuclear attack under deterrence is less than the chance of domination by the Soviets after US nuclear disarmament. He argues that although we do not have reliable quantitative estimates of the probabilities, the many methods of analyzing the nuclear situation that do assign numerical probabilities all seem to agree, at least in the ordering of their judgments, that a large nuclear attack would be a worse disaster than domination by an opposing power and that Soviet domination after US disarmament is more likely than a Soviet nuclear first strike under deterrence.9

According to DAP, if the greater disaster also has the greater probability, then it is easily decided that the action that avoids the greater chance of the greater disaster is the rational action. Yet with the structure of the problem that Kavka suggests, we are faced with a choice between a greater chance of a lesser disaster and a lesser chance of a greater disaster. Thus, when our choice must be made under two-dimensional uncertainty, but we are confident in our ordinal rankings of the values and probabilities of the

9Kavka, p. 45
potential disasters, then according to DAP, the rational action is *that action which minimizes the probability of disaster occurrence*. In this case, adoption of a deterrent policy is the rational action since it will minimize the probability of disaster occurrence.

Kavka believes that there are particular circumstances under which application of the DAP is most suitable. He suggests nine conditions that describe such circumstances. These are not axioms from which DAP is derived but rather are statements of limiting conditions for highly plausible applications of DAP:

1. The chooser lacks reliable quantitative probability and utility estimates.
2. The chooser has confidence in his orderings of the conditional probabilities of the various outcomes.
3. All disastrous outcomes are regarded as extremely unacceptable, (each involves large amounts of negative utility).
4. The disastrous outcomes are judged to be of roughly the same order of magnitude.
5. The chooser regards the utility disparity between the non-disastrous outcomes as being small compared to the utility difference between the disastrous and non-disastrous outcomes.
6. The choice is unique, i.e., not a series of like choices.
7. The probabilities of the disasters are not thought to be insignificant.
8. The probability of the greater disaster is not thought to be very large.
9. The probabilities of the disasters are not thought to be very close to equal.\(^{10}\)

There may be debate as to whether Kavka’s assumptions in setting out the problem meet these nine criteria. Kavka argues that they do. There also may be controversy over whether or not Kavka’s claims about the ordinal rankings of the probabilities and values of the potential disasters are true. One who raises such objections is Douglas Lackey, in his paper, "Missiles and

\(^{10}\)Kavka, p. 50-51
Morals: A Utilitarian Look at Nuclear Deterrence." He argues that Kavka's assumptions do not satisfy all nine criteria and that Kavka's ordinal judgments about values and probabilities of outcomes are wrong.

First, Lackey questions Kavka's claims that the disasters of nuclear attack and Soviet domination satisfy the fourth criterion, that the disasters must be of roughly the same order of magnitude. Lackey seems to want to argue that nuclear war would be orders of magnitude worse than domination, and thus, DAP fails to provide a credible decision.

To answer the objection, it must admitted that there is great difficulty in comparing values of nuclear war and Soviet domination. One possible way to estimate the value that a nation places on independence is to consider historical cases of nations fighting to prevent domination by a rival power. In WWII, the Soviets lost between 10 and 15 million soldiers and civilians as a direct result of their fight to stop the German invasion (others died later from starvation and disease). Despite these staggering casualties, the Soviets continued to resist the Germans. Thus, it could be suggested that a rough quantification of the value of Russian independence at that time was about 10-15 million lives.

Desire to preserve the independence of the US may be just as intense. Thus we might argue that there is some historical precedent for believing that US independence could be worth as much as 10 million lives. Lackey notes later in his paper that the estimates of US casualties from a Soviet first strike have ranged from about 800,000 expected dead to about 20 million expected dead. Although we can't determine the actual number before

---

12Lackey, p. 203

31
such an attack, a range of 10-20 million expected dead might be a reasonable estimate. Thus, we have a historical case of a nation willing to spend 10-15 million lives to preserve its independence and an estimate that perhaps as many as 10-20 million people in the US may die in a Soviet first strike. Thus, it seems plausible to argue that the potential disasters faced by the US could be of roughly the same order of magnitude.

Lackey also argues that Kavka's assumptions that the probability of nuclear war under deterrence is less than the probability of domination under nuclear disarmament are wrong.¹⁴ I don't think that there is a definitive answer to this objection. Liberal commentators on the subject of nuclear deterrence tend to rate the chance of war as high, while more conservative thinkers rate it as low. To make my point, however, I only need to assume that Kavka's claims about probabilities and values of outcomes are reasonably plausible and could be true; and I think this is the case. If we assume Kavka's claims to be true, then according to DAP, it is rational to adopt a policy of deterrence because it is the action that carries with it the lesser probability of disaster.

This conclusion that deterrence is the rational course of action is reached without appeal to claims about the value or probability of the act of retaliation. The rationality of adopting the deterrence policy is based on its lesser probability of resulting in disaster.

Let us now suppose that even though deterrence truly offered us the lowest probability of disaster, the Soviets launch a first strike in spite of our threats of retaliation. We now are faced with a new set of outcomes and

¹³Lackey attributes the figure of 800,000 dead to Secretary of Defense Arthur Schlesinger in Congressional testimony in 1974 and the 20 million figure to the US government's Office of Technology Assessment in 1979.
¹⁴Lackey, p. 204
probabilities to assess. As with MEU, it seems that the relevant choices of action are a choice between retaliating, or not retaliating.

In applying DAP to this decision, we might judge that retaliation carries with it the potential disaster that the nuclear destruction will be greatly increased to some unknown amount (our act of retaliation will double the nuclear devastation and perhaps will cause a Soviet counter-counter strike which will increase nuclear destruction in the US). The disaster for the US by not retaliating would be domination by the Soviets. Once the US chose not to respond to the Soviet nuclear strike, the Soviets would feel more confident in attempting domination or blackmail against the US since they would no longer fear retaliation. It seems plausible to suggest that greater nuclear devastation world wide and in the US would be a worse disaster for the US than would Soviet domination.

I think it's also reasonable to expect that the probability of increased nuclear devastation by retaliation is greater than the probability of domination after not retaliating. Once the US launches a retaliatory strike, I think there would be a high probability that the Soviets would launch their remaining missiles to prevent them from being destroyed. Thus, there would be a relatively high chance of further destruction to the US after retaliation. There is also a significant chance of Soviet domination if the US does not retaliate. Yet the chance that such domination could be successful is probably less than the chance of more destruction given retaliation. Assuming that these ordinal judgments are true, DAP tells us that retaliation is irrational because it provides for a greater chance of the greater disaster.
These examples suggest very plausible situations under which the DAP tells us that adoption of a deterrent policy is rational but that the act of retaliation would be irrational. Thus, we again see that it can be consistent with a widely held principle of rationality for it to be rational to intend an act that is irrational.

Maximin

The final theory of rationality that I want to consider is Maximin, which may be stated as follows:

Maximin: An action, A, is rational for subject, S, if and only if there are no alternative actions for S whose outcome is better than the worst outcome of A for S.

Maximin is often employed in cases where one cannot determine with confidence the probabilities that a given action will produce a given result. Under Maximin, one determines the worst possible outcome for each alternative action without attempting to determine the probability of the outcome. The rational action, according to Maximin, is then that action (or actions) whose worst outcome is the best.

Maximin does not show a preference between the choices of nuclear deterrence or unilateral nuclear disarmament. The worst possible outcome for the US, under a policy of deterrence, is that the Soviet Union ignores the threat of retaliation and launches a nuclear first strike that destroys the US. We also find that the worst possible outcome for the US under a policy of unilateral nuclear disarmament is that the Soviet Union launches a nuclear
first strike that destroys the US. Thus, the worst outcome is the same for either action.

Maximin tells us that the rational action is that action such that the worst outcome of any alternative action is no better than its worst outcome. If we have only two actions, and the worst outcome of each is the same (destruction of the US), then according to Maximin, both actions are rational.

Maximin does seem to be decisive when we consider the decision to be made if the Soviets were to launch a nuclear first strike. Our alternative actions would then be retaliation or no retaliation. The worst possible outcome for the US by choosing retaliation would be a Soviet launch of all their remaining nuclear missiles at the US once they detected launch of the US retaliation. This would result in large areas of the US being rendered uninhabitable with tens of millions of US citizens dead and millions more starving.

The worst possible outcome of not retaliating seems to be the similar case of the Soviets simply continuing to launch their missiles at the US until they have no more left. This outcome is perhaps very unlikely, given the fact that the US has not retaliated, but it would be the worst possible outcome.

The decisive factor between these two worst outcomes can be only that there would be fewer nuclear warheads exploded in the world under the action of no retaliation, because no US missiles would be fired. Nuclear effects throughout the world, such as fallout, nuclear winter, damage to food systems and oceans, etc. would be less under the action of no retaliation which anticipates fewer nuclear warheads exploding. It would be in the interests of the US to have less, rather than more, of the world contaminated
or destroyed by nuclear effects, so it seems that the action of no retaliation would be the rational choice.

Assuming that these claims about the worst possible outcomes are plausible, then it seems that it can be consistent according to the principle of Maximin for it to be rational to adopt a policy of deterrence and yet be irrational to retaliate.

Principles of Morality

We have seen that it can be consistent under common principles of rationality for the formation of the conditional intention to retaliate to be rational and the retaliation itself to be irrational according to the same principle. I now want to investigate whether we find the same result when we consider the morality of the action. Specifically, I want to consider whether there can be plausible circumstances under which it may be morally right to form the conditional intention to retaliate and for retaliation itself to be morally wrong according to the same principle of morality.

To do this, I will consider the principles of Maximization of Total Expected Utility and Disaster Avoidance, which hold that the morality of an action can be determined from the value and probability of the consequences expected from doing the action. These principles differ from those used to determine rationality in that they utilize measures of the total amount of utility that is expected to be produced by an action, rather than simply the expected utility for the agent doing the action.
Maximization of Total Expected Utility

The formulation of Maximization of Total Expected Utility (MTEU) is similar to Maximization of Expected Utility which was discussed previously except that rather than considering only the utility of a particular agent (such as the US) MTEU holds that the action which maximizes the total expected utility, is the action that is morally right. We can formulate MTEU as follows:

\[ \textit{MTEU}: \text{ An action is morally right if and only if it maximizes the total expected utility.} \]

An action, A, "maximizes the total expected utility" if there is no alternative action to A, with a greater balance of positive expected utility over negative expected utility than A has.\(^{15}\) To determine whether an action is morally right according to MTEU, we perform a calculation similar to that of MEU. First, one determines the possible outcomes from each alternative action open to the agent. We then determine the conditional probability (between 0-1) for each of these outcomes, such that the sum of the probabilities of all the outcomes of an action equals 1. We then assign a utility value to each of the outcomes. For each outcome of an action, we multiply its utility by its probability on the action. The sum of these products of the probabilities times the values for all of the outcomes of an action is the total expected utility for that action.

\(^{15}\)For my discussion, I will not attempt to define exactly what this measure of utility would be. The only distinction I will make is to note that there are positive and negative amounts of utility that correspond to consequences that are desirable or undesirable.
In the case we are considering, the agent is the US. The relevant actions that are open to the US remain the same as those of MEU: Deterrence or Nuclear disarmament. When we assign a utility value to each of the outcomes, we estimate the total amount of utility that is expected for the outcome, rather than just the utility to the US as in the calculation for rationality. The action that maximizes the total expected utility is morally right.

For this calculation, I'm suggesting that the relevant outcomes of deterrence are the same as we've considered before: Soviet first strike, Blackmail, Nuclear peace, and Accidental US strike, as well as the outcomes of the action of nuclear disarmament: Soviet first strike, Blackmail, and Nuclear peace. Probabilities of these outcomes on the action also remain the same. The expected utility values of these outcomes are different because they reflect the total expected utility rather than just the expected utility of the US. I believe that Table 2.3 shows a plausible assignment of values:

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Probability</th>
<th>Value</th>
<th>PxV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Soviet first strike</td>
<td>.1</td>
<td>-2000</td>
<td>-200</td>
</tr>
<tr>
<td>Blackmail</td>
<td>.1</td>
<td>-850</td>
<td>-85</td>
</tr>
<tr>
<td>Nuclear Peace</td>
<td>.75</td>
<td>2000</td>
<td>1500</td>
</tr>
<tr>
<td>US Accident</td>
<td>.05</td>
<td>-1000</td>
<td>-50</td>
</tr>
<tr>
<td><strong>Total:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Soviet first</td>
<td>.2</td>
<td>-2000</td>
<td>-400</td>
</tr>
<tr>
<td>Nuc. Disarm. Blackmail</td>
<td>.4</td>
<td>-850</td>
<td>-340</td>
</tr>
<tr>
<td>Nuclear Peace</td>
<td>.4</td>
<td>2000</td>
<td>800</td>
</tr>
<tr>
<td><strong>Total:</strong></td>
<td></td>
<td></td>
<td>+60</td>
</tr>
</tbody>
</table>
With this assignment of values, the value to the world of nuclear peace is as good as a nuclear attack by one nation on another is bad. The negative value assigned to Soviet blackmail includes the negative value to the US as well as the negative value to other nations (particularly other western democracies) that the Soviets may attempt to dominate or influence. Finally, the negative value of a US accidental strike includes the negative value to the Soviets as well as to adjacent countries that may be affected by the nuclear strikes. I'm assuming that an accidental US strike would be much smaller than a deliberate nuclear strike. Thus, the total negative expected utility of an accidental strike would be less than the total negative expected utility from a deliberate first strike.

As in the calculations for rationality, I only wish to claim that these probabilities and values are reasonably plausible and could be true. Assuming this, we find that adoption of a deterrence policy is the morally right action, according to MTEU, because it is the action that maximizes the total expected utility.

We can now consider the circumstances faced by the US if deterrence fails and the Soviets launch a nuclear first strike. The US then would be faced with the choice between retaliating and not retaliating. The relevant outcomes of retaliation are 1) more destruction— from the US retaliatory strike and potentially Soviet retaliation in response to the US retaliation. This would also include the value of attempted Soviet domination of the US. 2) End the war— US retaliation might destroy all the remaining Soviet missiles or serve to break the Soviet will to continue the conflict.

The relevant outcomes of no retaliation would be 1) End the war— if the US did not retaliate the Soviets might declare victory and end their
aggression. 2) Soviet domination- after a nuclear first strike to which the US did not respond, the Soviets could issue various economic, political and military demands throughout the world accompanied with the very believable threat of nuclear aggression if the demands are not met.

Assigning the same probabilities as I used in the earlier discussion of rationality, we have the following calculations (Table 2.4):

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Probability</th>
<th>Value</th>
<th>PxV</th>
</tr>
</thead>
<tbody>
<tr>
<td>More destruction</td>
<td>.55</td>
<td>-4000</td>
<td>-2200</td>
</tr>
<tr>
<td>Retaliation End the war</td>
<td>.45</td>
<td>+1000</td>
<td>+450</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td>-750</td>
</tr>
<tr>
<td>No retal. End the war</td>
<td>.6</td>
<td>+2000</td>
<td>+120</td>
</tr>
<tr>
<td>Soviet domination</td>
<td>.4</td>
<td>-1000</td>
<td>-400</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td>+800</td>
</tr>
</tbody>
</table>

I think that the outcome with the most damaging effect world-wide would be "more destruction," under the action of retaliation. This outcome anticipates a sequence of nuclear strikes between the US and the Soviets, where large portions of the world would be destroyed either directly from nuclear explosions or indirectly from fallout, nuclear winter, etc. These undesirable consequences would spread well beyond the borders of the US and Soviet Union. This outcome also includes the negative value of Soviet domination to the US of Soviet domination as well as the negative value to other nations that the Soviets may attempt to dominate after a nuclear exchange.

I have suggested a positive value of "end the war" under retaliation of +1000. This reflects the positive value of stopping Soviet aggression, but
since this also is accompanied by the effects US nuclear retaliation (particularly to adjacent nations that are not part of the conflict), this outcome has a lower value than does "end the war" under no retaliation. I have rated the outcome "end the war" under no retaliation higher than "end the war" under retaliation since the nuclear aggression would come to an end without further nuclear attacks.

We can see from the calculations in Table 2.4 that if the Soviets did launch a nuclear first strike, in spite of the US policy of deterrence, then the action of not retaliating would be morally right and retaliation would be morally wrong. Therefore, it seems that there can be plausible circumstances such that, according to the principle of maximization of total expected utility, it can be morally right to adopt a deterrent policy that includes the conditional intention to retaliate and be morally wrong to retaliate should the specified conditions occur.

Disaster Avoidance

The Disaster Avoidance Principle (DAP) that I presented earlier can easily be adapted to moral questions by considering the potential disasters in terms of the total expected negative utility they would produce rather than just the negative utility expected for the agent. In Kavka's presentation of DAP in "Deterrence, Utility and Rational Choice," he is concerned with the moral status of nuclear deterrence and considers the total amount of utility for each potential disaster in his application of DAP. We can formulate this version of DAP as follows:
DAP: when choosing between potential disasters under two-dimensional uncertainty, the alternative that minimizes the probability of disaster occurrence is morally right.

I will again consider that the alternative actions being considered are the policies of nuclear deterrence and nuclear disarmament. The potential disaster from the former being a Soviet nuclear first strike and the potential disaster from the latter being Soviet domination. As I've explained earlier, "two-dimensional uncertainty" refers to the claim that we do not have reliable estimates of the probabilities or the utility measurements for either of these disasters. There seems to be no natural reference class of past situations or scientific theories of international relations that could provide us with relative frequencies of nuclear war given nuclear deterrence of the type now practiced, or of the probability of Soviet domination given US nuclear disarmament in the present world situation.

Kavka does claim, however, that we can judge the ordinal ranking of the utilities and the probabilities of the disasters. We can judge that the death and destruction in the US, plus the nuclear effects to other nations, from a Soviet nuclear first strike would be a worse disaster in terms of total utility than would domination by the Soviets. Likewise, we can argue that the probability of a Soviet first strike under deterrence is less than the probability of Soviet domination given US nuclear disarmament. If these assumptions are correct, we are left with the decision between a lesser chance of a worse disaster and a greater chance of a lesser disaster.

Therefore, according to the DAP, adoption of a deterrence policy is the morally right action because it carries the lesser probability of disaster occurrence.
Let's now assume that despite the lesser probability of disaster, the Soviets launch a nuclear first strike. The US is now faced with a different set of alternative actions, namely, retaliation or no retaliation. It seems plausible to assume that the potential disaster from retaliation is repeated nuclear exchanges between the US and the Soviets from retaliation, counter-retaliation, counter-counter-retaliation, etc., until all the missiles have been fired. On the other hand, the potential disaster from not retaliating would be Soviet world domination. I think that we can judge that the potential disaster of all-out nuclear war is a worse disaster than Soviet domination in terms of total utility.

The probability that retaliation will add to the worldwide nuclear devastation is certain (unless all the missiles were duds), and the chance of additional Soviet strikes after US retaliation is probably very high. While the chance of Soviet domination after no retaliation is probably relatively high, it could be reasonably judged as less than the chance of greater nuclear devastation from retaliation. Thus, it seems reasonable to argue that the probability of the disaster of all-out war is greater than that of Soviet domination. We are left with a situation where the greater disaster also carries with it the greater probability, thus, the morally right action is no retaliation.

I'm not not making the claim that these ordinal judgments are true. My claim is that the judgments are made based on plausible combinations of circumstances and values that could be true. Thus, it can be consistent according to the Disaster Avoidance Principle for the conditional intention to retaliate to be morally right and the act of retaliation to be morally wrong.

At the beginning of this chapter, I suggested an argument with the premise "If retaliation is wrong (irrational) then the conditional intention
to retaliate is wrong (irrational). I then suggested that the potential consequences of forming the conditional intention to retaliate (e.g., prevention of nuclear war) may be sufficient for it to be right. Thus, the paradox: that the conditional intention to retaliate seems to be right and it seems to be wrong.

This apparent paradox forces us to examine more closely the reasoning that leads to these opposed conclusions about the normative status of the conditional intention to retaliate. In this chapter, I have concentrated on the line of reasoning that concludes that the conditional intention to retaliate can be right, based on the potential desirable consequences resulting from its formation. This is not a claim about the actual values and probabilities of the nuclear deterrence policy practiced by the US. Rather, it is the claim that there can be plausible circumstances such that the same principle will hold that the conditional intention to retaliate can be right while the act of retaliation is wrong. Whether the current world situation is an example of such a set of circumstances is a different question.

From the examples I have suggested, I think that we can identify a number of significant characteristics of the nuclear deterrence issue that come to light by applying these consequentialist principles:

1. It seems clear to me that nuclear deterrence, specifically the deterrence practiced by the US, must be characterized as primarily two separate actions. One action is the formation of the conditional intention to retaliate and the second action is the retaliation itself. This would be contrasted with a type of deterrence that was characterized by the activation of an automatic retaliation machine that would automatically launch a nuclear retaliation once it detected a nuclear attack, and that once turned on,
could never be turned off. Under this type of deterrence, only one action is required, i.e., turning on the machine. The action of turning on the machine includes both the deterrent intention to retaliate and the action of retaliating if attacked. The deterrence policy of the US, however, is not like this. US deterrence includes one action to adopt the policy, and then would require a second, separate action of retaliation.

2. The actions of formation of the conditional intention and the action of retaliation each have independent utility values. In other words, the utility value associated with forming the conditional intention to retaliate does not derive from or include any positive or negative utility values of retaliation.

3. Each of the two actions has a unique set of outcomes. These different outcomes reflect the different purposes and intentions for doing each action. As we saw in the examples in this chapter, the desired result, or intention, of forming the conditional intention to retaliate is prevention of a Soviet first strike. On the other hand, the action considered if deterrence fails (retaliation) will be for the purpose of ending the war, limiting nuclear damage, or preventing Soviet world domination.

4. It is clear that within the framework of each of the consequentialist principles that I have considered, it can be consistent for the conditional intention to do an act to have a normative status that is different from the normative status of the act itself. It's interesting to note that when we apply purely consequentialist principles to the case of nuclear deterrence, and get results where an intention to do an act is right, but the act itself is wrong, we do not perceive that a paradox arises. I suspect that this is due to the points that I have listed above. By viewing the action of formation of the conditional intention as a separate act from the act of retaliation, and by
realizing that the formation of the intention carries with it its own utility value, we see no problem in situations where, according to consequentialist principles it may be right to intend to do an act that is wrong.

5. I think that the consequentialist principles that I have discussed in this chapter fail to address whether there is some type of special relationship between an action and the intention to do the action. The argument that I presented at the beginning of this chapter relied on just such a relationship, namely that if an action is wrong then the intention to do the action is also wrong, regardless of its consequences.

From the examples in this chapter, it seems that the line of reasoning, based on consequentialist principles, that concludes that the conditional intention to retaliate can be right even when the retaliation itself is wrong is plausible. Thus, we must now examine the line of reasoning, based on relevant non-consequentialist principles, that suggests that if retaliation is wrong then the conditional intention to retaliate must also be wrong. This will be my project in Chapter 3.
CHAPTER 3

NON-CONSEQUENTIALIST PRINCIPLES

We saw in chapter 2 that one of the lines of reasoning that leads to paradox - the line of reasoning that concludes that the conditional intention to retaliate is right based on its consequences, seems plausible (I refer to this as the consequentialist line of reasoning). It seems that there can be reasonable circumstances under which a consequentialist principle will hold that the conditional intention to retaliate is morally right (rational) and the act of retaliation is morally wrong (irrational). Yet this result depends importantly on the claim that formation of the conditional intention to retaliate and the retaliation are two independent actions. Because we view the issue as an evaluation of two independent actions, we assume that there will be two separate utility values. Thus, there is no conflict in a situation where one action is evaluated as right and the other action as wrong.

There is, however, a second line of reasoning about this conditional intention that conflicts with the consequentialist account to form the paradox. The reasoning here depends on appeal to moral principles which hold that the action of retaliation, and the action of forming the conditional intention to retaliate are not independent actions, but are related in an important way. This line of reasoning is non-consequentialist in that the moral principles assumed, and according to which the normative statuses of actions are determined, appeal to features of the actions other than their consequences. In the case of nuclear deterrence, the action of nuclear retaliation is determined to be morally wrong according to a specified, non-consequentialist moral principle. The conditional intention to retaliate is
then determined to be morally wrong by appeal to a non-consequentialist moral principle which holds that it is wrong to intend (even conditionally) to do an action that is itself wrong. In both cases, the normative statuses of the actions are determined according to moral principles that appeal to features of actions other than their consequences. (I refer to this as the non-consequentialist line of reasoning).

Based on these moral principles, and some assumptions about nuclear retaliation, the non-consequentialist line of reasoning concludes that nuclear retaliation is morally wrong and therefore, the conditional intention to retaliate is also morally wrong. In this chapter, I want to examine more closely the principles, assumptions and arguments required for this line of reasoning to work, in order to see whether there is the same degree of plausibility in application of this line of reasoning as we had in the application of the consequentialist principles.

A typical non-consequentialist evaluation of nuclear deterrence includes two moral principles and one assumption. These are:

**The Principle of Discrimination:** The deliberate killing of non-combatants is morally wrong.

**The Wrongful Intentions Principle (WIP):** It is morally wrong to intend what it is morally wrong to do.

**Assumption:** Nuclear retaliation requires the deliberate killing of non-combatants.

With these principles and the assumption, an argument can be formulated to show that nuclear retaliation is morally wrong. Such an argument might be as follows:
**Argument A1:**

1. The deliberate killing of non-combatants is morally wrong
2. Nuclear retaliation requires the deliberate killing of non-combatants
3. If an act, A, is morally wrong, and another act, B, requires A, then B is also morally wrong.
4. Therefore, nuclear retaliation is morally wrong.
5. It is morally wrong to intend (even conditionally) to do what is itself morally wrong.
6. Therefore, the conditional intention to retaliate is morally wrong.

If we were to add a premise such as, "Nuclear deterrence is the conditional intention to retaliate," then we could also draw the conclusion that nuclear deterrence is morally wrong. The argument is valid, but we must examine the truth of the premises to determine its soundness. Before I turn to discussion of the premises, I first want to note some other versions of this argument.

One variation of A1 can be formulated by using the Principle of Proportionality in place of the Principle of Discrimination. The Principle of Proportionality asserts that a military action would be morally wrong if the number of innocent civilians killed by the action exceeded what would be proportional to the military value of the action. The reasoning of this version of A1 is that the great number of civilians that would be killed as a result of the use of nuclear weapons against a particular target will always exceed what would be proportional to the military value of that target.

Thus, premise 1 of A1 is replaced by "1. A military action is morally wrong if the number of innocent civilians killed by the action exceeds what would be proportional to the military value of the action." and premise 2 is replaced by, "2. The number of innocent civilians killed by nuclear
retaliation exceeds the number that would be proportional to the military value of nuclear retaliation." The rest of the argument remains the same, resulting in the conclusions that the conditional intention to retaliate is morally wrong and therefore, nuclear deterrence is morally wrong.

There is another variation of the non-consequentialist argument that I also want to mention. Instead of the WIP, this argument uses the Wrongful Threats Principle (WTP): *That it is wrong to threaten to do an act that is itself wrong*. This principle has been defended by David Hoekema. He argues against Kavka's discussion of intentions, noting that,

"...Kavka fails to recognize that the WIP does not bear as directly as he supposes on deterrent situations. For it is not the intentions that deter, as Kavka supposes, but threats."16

"In the characteristic case of conditional threats, one person threatens another because she wants him to act in a certain way. What has this effect, if the threat succeeds, is the threat itself-- the declared intention to do harm."17

I think that Hoekema is correct in claiming that it is threats, rather than simple intentions, that have deterrent effect. However, he then makes the claim that,

"A threat is the declaration of an intention to do harm, and in order to be effective the declaration must be credible. But whether the threatener actually holds such an intention is irrelevant to the effectiveness of the threat."18

---

17 Hoekema, p. 100
18 Hoekema, p. 100
I think that this claim does not hold in the case of nuclear deterrence. Kemp has noted that a deterrent threat could be accompanied by any of three possible intentional states. First, the threat to do an act could be accompanied by the conditional intention to do the act if the threat is unsuccessful in deterring the unwanted action. Second, the threat could be accompanied by a secret intention not to do the act if the threat is unsuccessful in deterring the unwanted act. Thus, the threat would be a bluff. Finally, the threat might not be accompanied by any formed intention at all. The agent simply might not know what he would do if the threat is unsuccessful in deterring the unwanted action.19

As I've discussed in Chapter 1, given the organization of the US military and the US industrial base, I don't think that the US could keep secret the fact that the threat was a bluff. If it became known that the threat to retaliate was a bluff, then clearly it would have no deterrent value.

Likewise, it has been claimed that for the deterrent to be effective, we must convince the opponent that there is no question that retaliation will take place and that it will result in unacceptable consequences.20 In order for this to be accomplished, the opponent must believe that the agent will retaliate, and has the capability to retaliate effectively. This requires that the agent develop and stockpile appropriate weapons systems, develop procedures and strategies for their use, select targets, and train personnel to use them. I don't think that these tasks could be accomplished, particularly in a manner that would effectively deter an opponent, if the agent did not form the sincere intention to retaliate.

19 Kemp, p. 277-8
20 Secretary of Defense Weinberger, “To maintain a sound deterrent, we must make clear to our adversary that we would decisively and effectively answer his attack,” Annual Report to Congress, FY 1984, February 1, 1983
Thus, in the case of nuclear deterrence, I want to claim that for the threat to retaliate to be credible and effective, it must be accompanied by the sincere intention to retaliate. Contrary to Hoekema's claim, I think that whether or not the threatener forms the sincere intention to retaliate is vitally important to the effectiveness of the threat.

It should also be noted that if these threats would not successfully deter the Soviet Union, then it is unlikely that either threat would have the best expected consequences of the alternative actions available to prevent nuclear attack. Therefore, neither would be morally right according to the consequentialist line of reasoning, and we find that the paradox fades away.

For nuclear deterrence to be credible and effective, and for paradoxical deterrence to arise, the threat to retaliate must be accompanied by the sincere intention to retaliate. Likewise, the sincere intention to retaliate must be expressed as a declared intention, or threat, in order to influence an opponent. I am more interested in the implications of forming the intention to retaliate, and how such an intention leads to paradoxical deterrence. Thus, instead of threats I will concentrate on the intention to retaliate and the argument that utilizes the WIP, as representative of the non-consequentialist line of reasoning (and will assume that this intention is expressed as a threat). I suspect that my discussion will apply similarly to threats and arguments utilizing the WTP.

An objection might be raised to premise 2, of A1, that retaliation requires the deliberate killing of innocents, by claiming that nuclear retaliation is an attack against legitimate military targets. Any innocent civilians that are killed is a result of unintended effects of the legitimate military attack, and thus, retaliation does not require the deliberate killing of innocents. I think
that there are two responses to this objection. One response is to accept the objection to premise 2, but then offer the version of A1 that utilizes the principle of proportionality that I've discussed above. In this version of A1, even nuclear weapons fired at specific, legitimate, military targets will be morally wrong.

Another response to this objection is to point out that nuclear retaliation does indeed require the deliberate killing of innocent civilians. It is the very fact that innocent civilians will be killed, and the vast number that will be killed, that is supposed to have the deterrent effect on the opposing government. Therefore, I think that we can accept premise 2 of A1.

There seems to be much intuitive support for arguments such as A1 that utilize the Wrongful Intentions Principle. We can think of many cases where it seems true that it is wrong to intend to do an act that itself is wrong. Yet, I think that there are some interesting claims that suggest that the conclusions of A1, and the applicability of the WIP, are not quite so obvious. I want to present two such arguments.

One argument, suggested by Jonathan Schonsheck,\(^2\) acknowledges that A1 can be used to argue against nuclear deterrence. He claims, however, that although formation of the intention to retaliate is morally wrong according to A1 (based primarily on application of the WIP), the argument does not show conclusively that the intention to retaliate is morally wrong. His argument suggests that the moral evaluation rendered by A1 may be overridden by other considerations or other moral principles that may apply.

I then want to present an argument by Kavka for the claim that there are peculiar characteristics in the case of nuclear deterrence such that the WIP should not be applied in the normative evaluation of the conditional intention to retaliate. Once these characteristics are identified, it becomes clear that the factors that made WIP applicable to most intentions, and that make it intuitively acceptable, are not present in the case of nuclear deterrence.

Schonsheck does not argue specifically for the moral permissibility of nuclear deterrence, but rather against the immediate rejection of deterrence based on the WIP and the claim that retaliation is morally wrong. He suggests that we must specify more carefully the role that principles such as the WIP play in making judgments about actions.

Schonscheck suggests that there are two "understandings" of the role that non-consequentialist principles play in the moral assessments of actions. First, is what he calls the "conclusive understanding." According to this understanding, "a finding that some action would violate some non-consequentialist moral principle is sufficient for the action's being wrong, for its being morally impermissible."[22]

For example, arguments such as A1 make the case that retaliation is wrong, and then by appeal to the WIP, conclude that the conditional intention to retaliate is morally wrong. These arguments suggest that the WIP, and the premise that retaliation is wrong, comprise a decisive case. They hold that the WIP is an absolute principle, and thus, forming the conditional intention is wrong regardless of the consequences and regardless of any other moral principles that may apply to the action.

[22]Schonsheck, p. 333
He then contrasts the "conclusive understanding" with what he calls the 
Prima Facie understanding. According to this understanding,

"that an action would violate some non-consequentialist principle 
counts against that action, constitutes a moral impediment to its 
permissibility. However, that the action would violate such a principle does not establish conclusively that the action is wrong, that it is morally 
impermissible for a person to do it."23

Schonsheck suggests a number of reasons that we may prefer the 
"Prima Facie" understanding to the "conclusive" understanding. It may be 
that in the particular moral choice situation that a person finds himself, 
every available alternative may violate a non-consequentialist principle. If 
some alternative must be selected (and even not selecting one of the 
available alternatives would violate a non-consequentialist principle), then 
some principle must be violated. In such a situation, it seems that the only 
recourse is to claim that some violations must be better or worse than others. 
If this is the case, it would be premature to condemn an action for its being a 
principle violation before we examined what principles would be violated, 
and how they would be violated, by the available alternatives to the action.

We might also think of this in terms of duties. Thus, if there is a moral 
principle that specifies that a person should do action A under certain 
morally relevant circumstances, then we could say that the person has a duty 
to do action A under those circumstances. But can we conclude from this that 
action A is the action that the person ought to do? It seems not. For one can 
imagine a variety of circumstances where more than one moral principle 
may apply in a given situation, creating conflicting duties. It may be that 
one duty can be fulfilled only by neglecting another.24 Under these

23Schonsheck, p. 333
circumstances, it seems that one must determine which duty is more important, or more pressing, and act to fulfill that duty.

Once a person has found himself in a situation of conflicting duties, and chosen to act in accordance with the duty determined to be the most important, Schonsheck asks what can we say about the choice. He notes that we can point to the duty not fulfilled and conclude that the person failed to satisfy some duty. Or we can look to the duty that was fulfilled and the person's reasoning for determining that that duty was the one that should be followed. Schonsheck argues that,

"If we agree that the moral agent in question did in fact find himself or herself in a situation of conflicting duties, it is incomplete and morally untoward to merely point out the duty unfulfilled, neglecting the mention of the duty that was in fact fulfilled, and the reasons of the agent for believing that that was the more important duty." 25

One can certainly question and debate the reasons and evidence the moral agent used for determining which was the more important duty. But to simply point to a duty in isolation, note that it was not fulfilled, and then claim that the action was wrong, seems to be an incomplete moral evaluation of the action and the agent.

Schonsheck uses terminology suggested by W.D. Ross to talk about these situations of conflicting duties. He suggests that,

---

24 Schonsheck, p. 334
25 Schonsheck, p. 334
"When a moral rule of the duty-specifying sort applies to a person, that person is said to have a *prima facie* duty to perform the action specified by that rule....But while a reason, it is not in isolation a *conclusive* reason. There may be another duty-specifying moral rule which applies to the person in his/her circumstances; it too yields a *prima facie* duty. The right-acting person fulfills the more pressing duty.... The more important duty is called one's duty *all things considered*. A person in these circumstances is not culpable, or less virtuous, for having left the other (*ex hypothesi*) less pressing duty unfulfilled."\(^2^6\)

He also suggests that we can use this terminology of "*prima facie*" duty and duty "*all things considered*" *mutatis mutandis* as regards "wrongs." Thus, an agent may be in a situation such that each available alternative is a violation of some moral principle, i.e., each alternative is *prima facie* morally wrong. In such a situation, it seems that since all the alternatives are "wrong," the right thing to do is to choose that action that is "least wrong." Thus, what the agent ought to choose will be *prima facie* wrong, but will be the right thing to do, *all things considered*.

Schonsheck notes that,

"Whatever is selected will still be wrong in the *prima facie* sense; that it is a rule violation does not change, and there is still moral reason against its selection. However, it is true (*ex hypothesi*) that the moral case against the other alternatives is stronger, and it that fact that justifies the selection."\(^2^7\)

Thus, one immediately apparent difficulty with the non-consequentialist line of reasoning that relies on A1, is that it focuses solely on the WIP to determine the normative status of the conditional intention to retaliate. It presents an argument to show that retaliation is morally wrong, and then argues that formation of the conditional intention to retaliate is also morally

--

\(^2^6\) Schonsheck, p. 334

\(^2^7\) Schonsheck, p. 335
wrong because it violates the WIP. Yet it does not present relevant arguments to show that other non-consequentialist principles, that may apply to the conditional intention, are less pressing and should be passed over in favor of the WIP.

Objections to Schonsheck's argument might take a number of forms. For example, if one's moral system consisted solely of the Wrongful Intentions Principle, then there would be no concern about application of other moral principles to the conditional intention. One would simply apply the WIP, so to speak, and get a normative evaluation of the conditional intention. Yet such a moral system is clearly absurd, for one would not even be able to determine the normative status of the act itself, since the WIP only makes a claim about intentions.

One might also claim that the WIP was the preeminent principle of their moral system. Thus, there would be no other moral principle in the system that could override the evaluation rendered by the WIP. If someone subscribed to such a system, then once it was determined that the conditional intention was wrong according to the WIP, no other principle would have to be applied to the intention since it is stipulated that no other principle would override the WIP. This again seems to be a rather absurd moral system.

Finally, it might be claimed that one's moral system was so complete and consistent that conflicts among its principles were not possible. Thus, once an action was determined to violate one of the principles, it could be rejected without consideration of any other principles in the system.

Unless one subscribes to a moral system of one of these types, it seems possible that an agent may find himself or herself in a situation where he or she has to choose between alternatives each of which violates some non-
consequentialist principle. If this is true, then it seems that rejecting any of the alternatives prior to a full evaluation of all the alternatives is premature and unwarranted.

Anthony Kenny subscribes to a moral system of divine law, in which presumably there could not be conflicts of duty. In this system, were one to determine that an action violated a divine law, then one can immediately reject the action as wrong. Kenny argues that, "One of the actions prohibited by divine law is the intentional killing of the innocent." 28 He then notes that the,

"...NATO defence policy involves a readiness to commit murder on a gigantic scale. The intention to do so is admittedly a conditional one. But one may not intend even conditionally to do what is forbidden absolutely."

"We must give up our nuclear deterrent not because by so doing we shall achieve some desirable aim, but because to retain it is wicked. What will then follow is not in our hands. The prospect of standing defenceless before Communist Russia is indeed a sombre one. But that does not justify us in covenanting with the NATO powers to commit murder." 29

Kenny is using two important principles in his argument that nuclear deterrence should be rejected. First is a divine law that killing innocents is absolutely forbidden. Second, is the WIP. These principles, with the assumption that nuclear retaliation is killing of the innocent, yield the conclusion that nuclear deterrence is absolutely forbidden.

Kenny does not directly address the distinction between "prima facie" wrongs and "all things considered" wrongs. He suggests in the passages

29Kenny, p. 72
30Kenny, p. 74
above that once it is established that an action violates a divine law, then it can be rejected. It seems that Kenny also must hold that it is not possible within his system for there to be a conflict between divine laws. Thus, he must assent to the claim that absolutely nothing could justify "covenanted to commit murder."

It seems possible, however, to imagine cases that show conflicts between divine laws, or at least what humans might perceive to be conflicts between divine laws. In such a case, it would still seem warranted to consider the other available alternatives, since our perception that the act we are about to reject violates a divine law could also be confused. Even after such consideration, it might still appear that there are conflicts between the divine laws in a given situation.

For example, suppose one has a nuclear deterrent policy in place. Suppose it's also the case that the conditional intention to retaliate is the best and most effective means with which to keep a hostile, nuclear armed opponent from attacking. To switch to a policy of nuclear disarmament will bring about a nuclear attack on the agent's nation, killing millions of innocent citizens. It seems that the action of nuclear disarmament is absolutely forbidden. But according to Kenny's argument, continuing the policy of deterrence is also forbidden. Kenny's argument does not address how we should choose between these undesirable alternatives.

I don't wish to debate the moral system of divine law. But it strikes me that it is possible to conceive of situations where there seem to be conflicts among divine laws, or at least where we perceive the divine laws to yield conflicting results. In such cases, it seems more reasonable to consider all the available alternatives, and select the one that seems "least wrong", rather
than pointing to a principle or divine law in isolation, note that the action violates this one law, and then reject the action.

It's also interesting to note that Kenny does not provide any arguments in support of the WIP. He does assert that killing the innocent violates divine law. But he provides no justification for the claim that if an act is wrong, then intending to do the act must also wrong. Although this is a crucial premise in the evaluation of the conditional intention, he does not argue for the principle.

Kenny's divine law moral system may be consistent enough so that there cannot be conflicts between divine laws. Yet many philosophers who cite arguments such as A1 are not arguing from a divine law moral system. Thus, Schonsheck's argument that it is premature to dismiss the conditional intention to retaliate without a full appraisal of all the available alternatives is important.

I think that Schonsheck's claim, that one should pursue an evaluation of the conditional intention to retaliate, all things considered, rather than rejecting it once we determine it is prima facie wrong is plausible. Thus, I think that we have a potentially important objection to the non-consequentialist line of reasoning that rejects the conditional intention to retaliate. This is not to conclude that the conditional intention is morally right. Rather, it is the claim that A1 is not sufficient to declare that the conditional intention to retaliate is morally wrong. A1 would have to be paired with further arguments to show that the alternatives to nuclear deterrence were "less wrong."

A different type of objection against A1 concentrates on rejecting premise 4, the Wrongful Intentions Principle. This is the premise that is
particularly relevant to paradoxical deterrence, thus, for the sake of my discussion, I will assume that retaliation may be morally wrong (based on lines 1-3 of A1).\textsuperscript{31} What I am interested in is whether we can make the move at premise 4, of A1, which leads to the conclusion that the conditional intention to retaliate is morally wrong solely on the basis of the act of retaliation being morally wrong. This is the premise that is required to make the non-consequentialist argument work, but it is, I think, open to serious objection.

Arguments such as A1 seem persuasive primarily because the WIP seems obviously true. Gregory Kavka has noted that, "WIP seems so obvious that, although philosophers never call it into question, they rarely bother to assert it or argue for it."\textsuperscript{32} It also seems that the WIP may apply to conditional intentions in the same manner as it does non-conditional intentions. Kavka has noted that,

"Suppose I form the intention to kill my neighbor if he insults me again, and fail to kill him only because, fortuitously, he refrains from doing so. I am as bad, or nearly as bad, as if he had insulted me and I killed him. My failure to perform the act no more erases the wrongness of my intention, than my neighbor's dropping dead as I load my gun would negate the wrongness of the simple intention to kill him."\textsuperscript{33}

But why does the WIP seem true? It may be that many people associate the intention to do an act with the desire to do the act. Thus, one might claim that as a desire to do evil is evil, so the intention to do evil is also evil.

\textsuperscript{31}Retaliation, under certain circumstances, might be morally right. If it were, however, there would be no paradox. Thus, the philosophically interesting cases are those where retaliation is morally wrong, but the status of the intention to retaliate is in question.

\textsuperscript{32}Kavka, Some Paradoxes of Deterrence", p. 289

\textsuperscript{33}Kavka, p. 289
Similarly, many people may think that by forming the intention to do evil, one is taking actions to make the evil act come about. This also seems to make the intention evil. Some people may also think that the formation of the intention is in some way the start of the act itself. For example, Kavka suggests that one reason people may regard the WIP as true is that,

"...it is convenient, for many purposes, to treat a prior intention to perform an act, as the beginning of the act itself. Hence we are inclined to view intentions as parts of actions and to ascribe to each intention the moral status ascribed to the act "containing" it."34

It's unclear whether Kavka actually holds this view or whether he is simply suggesting it as an explanation of why some people may believe the WIP to be true. In either case the view seems wrong. In the passage above, he seems to be suggesting two claims. First, he is suggesting that the intention to do an act should be considered as the beginning part of the act itself. If this claim were true, then in the case of nuclear deterrence, we would have to say that when the conditional intention to retaliate which was formed almost 40 years ago, the act of retaliation was begun, and apparently has been in progress ever since. This seems to be a rather absurd claim.

This claim seems particularly implausible in the case of conditional intentions, where a conditional intention is formed at one time, but the act occurs, if it ever occurs, at a much later time. Likewise, we can think of many sincere intentions that are formed but where the act never occurs. Either the intention is abandoned before the act occurs, or, as in the case of conditional intentions, the antecedent condition of the conditional intention just never obtains. Thus, it seems incorrect to claim that the intention to do an act is the beginning of the act.

34Kavka, p. 289
The second claim suggested by Kavka's remarks is that, as a "part" of an evil act, the intention to do the act must also be evil. There seem to be many cases where this does not hold. A murderer's act of murder would then have to include as a "part" the intention to murder as well as other "parts" such as driving his car to the murder site, obeying traffic laws, walking down a street, etc. If this second claim were true, then each of these "parts" must be labeled morally wrong if the act of murder is wrong. But it seems absurd in this case to claim that the acts of driving a car and obeying traffic laws are morally wrong. Thus it seems that both claims, that the intention to do an act is the beginning that act and that the intention must have the same normative status as the act of which it is a "part," are implausible.

The confusion in these cases may stem from the fact that in most cases the only noticeable effects of the intention to do an act is the occurrence of the act itself (and the consequences of the act). This may cause some people to accept that the intention is part of the act or should be given the same normative status as the act.

Many of these reasons that make the WIP seem true are not present in the case of nuclear deterrence. Kavka and others\(^{35}\) have argued that the WIP should not be applied to the case of nuclear deterrence because the conditional intention to retaliate present in nuclear deterrence policies has such unique characteristics that the WIP is not applicable to it. Thus, Kavka rejects premise 4 of A1.

As I've noted, WIP seems intuitively appealing because we usually think that when an agent forms an intention to do an act, even a conditional intention, the agent desires to do the act. We may also believe that the purpose of forming the intention is to make the action come about. Finally, it may be that the only consequence of forming the intention is the bringing about of the act itself. In these cases, little importance attaches to the distinction between the intention and the intention of the intention.\footnote{Kavka, "Some Paradoxes of Deterrence," p. 291}

Kavka has argued, however, that this is not the case with the conditional intention to retaliate in nuclear deterrence. The ground of the desire to form the intention is entirely distinct from any desire to carry out the intention. The purpose, or intention, of the conditional intention is as a means of deterrence, not as the first step in carrying out the act of retaliation. Kavka argues,

"He desires \textit{having the intention} as a means of deterrence. Also, he is willing, in order to prevent the offense, to accept a certain \textit{risk} that, in the end, he will apply the sanction. But this is entirely consistent with his having a strong desire not to apply the sanction, and no desire at all to apply it. Thus, while the object of his deterrent intention might be an evil act, it does not follow that, in desiring to adopt that intention, he desires to do evil, either as an end or as a means."\footnote{Kavka, p. 291}

This argument points to the fact that there are effects from the conditional intention that are distinct from the effects of the act intended. In most cases, the only significant effects of an intention that can be identified are those of the doing of the action and the consequences of the action. Thus, if evil actions are intended that would produce evil consequences, the only
noticeable effects of forming the intention are the evil action itself and the evil consequences.

Again we find that this is not the case with the conditional intention to retaliate. The agent believes that forming the conditional intention will have a particular and significant effect, namely, prevention of a nuclear first strike by his opponent. This effect is completely independent of any effect that would result from retaliation. Kavka has called these the "autonomous effects"\(^{38}\) of the deterrent intention. The autonomous effects of the conditional intention to retaliate are specifically the influence that having the intention will have on the agent's opponent's behavior. It seems that these autonomous effects are relevant to the normative evaluation of the formation of the intention. The WIP, however, evaluates the intention solely on the effects of doing the action and the consequences of doing the action, while ignoring the effects of the forming of the intention itself.

The interesting, and unusual, features of these autonomous effects in the case of the conditional intention to retaliate is the belief that the autonomous effects of forming the intention will insure that the action of retaliation will never happen. The conditional intention to retaliate is formed in the belief that it will prevent the antecedent condition of the conditional intention from occurring. Normally we think that an intention initiates a sequence of events that have as their expected result the performance of the action that was intended. But in the case of nuclear deterrence, the conditional intention is formed in the belief that it will prevent performance of the action that was intended. Thus, we might describe the

\(^{38}\text{Kavka, p. 291}\)
conditional intention to retaliate as a self-defeating intention; the effects of the intention will make it unnecessary to do the action intended.

In forming the conditional intention, the agent believes that it will influence his opponent's behavior (prevent him from launching a first strike) and thus, will prevent the antecedent condition of the conditional from ever obtaining. With these points in mind, it's misguided to assess the normative status of the intention to do an act solely in terms of the normative status of the act. Rather, it seems that the normative evaluation should be made of the conditional intention as an act in itself. The WIP, however, attaches the normative status of the act to the intention to do the act.

Kemp cites an objection offered by John Finnis to the claim that this self-defeating characteristic is relevant to the evaluation of the conditional intention. Finnis supposes that a group of bank robbers, fearing the consequences of committing homicide, might nonetheless decide to shoot if their robbery is interfered with, and may even judge that the chances of any interference will be minimized if they make threats with their guns. They may believe that these threats will have the effect of influencing the bank employees' behavior such that the antecedent condition of their threat ("if anyone tries to stop us") will never be realized. But they still intend to shoot if their threat fails to prevent interference in their robbery. The claim here is that these threats, and their associated conditional intentions, are still wrong even though they have similar characteristics as are claimed for nuclear deterrence.

---

Kemp notes, however, that there are some important differences. First, the bank robbers are using the threats as a means to completion of an evil project, namely the robbing of a bank. Whatever the assessment of nuclear deterrence itself, it seems uncontroversial that its goal, or project, specifically the protection of the Western democracies from nuclear attack, is at least a *prima facie* moral imperative.\textsuperscript{40} Thus, we have good reason to condemn the bank robbers.

The second difference is that the conditionally intended killing plays different roles in the two cases. The bank robbers' project is to get away with the money. Actual killing may be necessary to achieve that end. If their deterrent fails, performance of the intended killing will still contribute to the successful completion of their project. Their choice is either to kill to achieve their goal or to surrender. In the case of the agent practicing deterrence, however, the conditionally intended killing is not similarly related to his end or goal. His goal is the prevention of a nuclear first strike. Once a first strike has occurred, performing the conditionally intended killing does not serve to further his goal of preventing a nuclear first strike. He has already failed in achieving his goal. It was the intention alone that promoted his end, through the autonomous effects he believed it to have on his opponent. Kemp argues that,

"Given that the deaths of innocents (as opposed to the intention to kill them) serves no purpose of theirs [those who practice nuclear deterrence], and is not chosen as a means to some further goal at which they aim, it is not possible to say even about the maintainers of anti-city deterrent the things that Finnis & co. say about other bearers of an intention to do what is immoral."\textsuperscript{41}

\textsuperscript{40}Kemp, p. 291

\textsuperscript{41}Kemp, p. 292
In the case of most intentions it might seem pointless to adopt an intention that is self-defeating because in most intentions the agent desires to perform the intended action. The conditional intention included in nuclear deterrence seems to be an exception. It's not pointless because we desire the autonomous effects of having the intention. This makes it worth forming the conditional intention to retaliate even if we believe that it will insure that the object of the intention is never performed.

It seems to me that when we examine the conditional intention to retaliate more closely, we find that many of the characteristics commonly associated with intentions, characteristics that make the WIP intuitively appealing, are not present. In the case of nuclear deterrence, there is no desire to do the act of retaliation. The formation of the conditional intention creates effects that are independent of doing the action intended and the consequences that would result from the doing the action. These independent effects are of immense value (prevention of nuclear war). Finally, the conditional intention is self-defeating; if it is successful, the object of the intention will never be performed. If it is unsuccessful, and a first strike occurs, it does not further the agent's project to do the retaliation. Because the intention behind forming the conditional intention is so completely oriented towards the effects of having the intention, and not with desiring the object of the intention (retaliation), I think it is misguided to evaluate the normative status of forming the conditional intention solely on the normative status of retaliation itself.

All of these factors combine to make application of the WIP to nuclear deterrence very implausible. The non-consequentialist line of reasoning may still hold that retaliation is morally wrong because it violates absolute
principles such as the Principle of Discrimination or the Principle of Proportionality. But the non-consequentialist line of reasoning that rejects forming the conditional intention to retaliate also requires the WIP as a key premise in its argument. If the WIP is shown to be not true, or shown not to apply to intentions that are part of a nuclear deterrent policy, then the non-consequentialist argument fails and no paradox arises.

I think that the non-consequentialist line of reasoning is very questionable due to its reliance on the WIP in its argument against nuclear deterrence. I don't know if the unique characteristics of the conditional intention to retaliate demonstrate conclusively that the WIP is not true and should be abandoned completely. Yet I think that it does suggest that there are cases where the WIP should not be applied because it fails to account for all of the relevant moral factors required to make an accurate and plausible normative evaluation. The conditional intention to retaliate that is part of a nuclear deterrent policy is one such case.
My purpose in this thesis has been to focus on a paradox found in some types of nuclear deterrence. The paradox arises specifically in the normative evaluation of the act of forming the conditional intention to retaliate. There are two lines of reasoning about the normative status of forming the conditional intention to retaliate that conflict to form the paradox. One line of reasoning holds that the consequences of forming the conditional intention to retaliate (e.g., prevention of nuclear war and nuclear blackmail) make formation of the intention morally right. A second line of reasoning, however, holds that retaliation itself is morally wrong, and thus, the conditional intention to retaliate must also be morally wrong. I have labeled deterrence, in which these two lines of reasoning conflict, as "paradoxical deterrence."

Each of these lines of reasoning requires certain assumptions about nuclear deterrence. The first line of reasoning (consequentialist) holds that forming the conditional intention is morally right because it is the best way to prevent nuclear attack (it has the best consequences). Thus, for the consequentialist line of reasoning to be successful, and for the paradox to arise, we must assume that forming the conditional intention to retaliate is the best way to prevent nuclear attack. One could make the claim that forming the conditional intention to retaliate is not the best way to prevent nuclear attack. Yet if some other policy could more effectively prevent a nuclear first strike, then that policy would be morally right, rather than a
policy of nuclear deterrence. In such a case, the paradox would not arise because the consequentialist line of reasoning would no longer hold that forming the conditional intention to retaliate was morally right.

The second line of reasoning (non-consequentialist) holds that nuclear retaliation is morally wrong, and therefore, forming the intention to retaliate (even the conditional intention) is also morally wrong. Thus, for the non-consequentialist line of reasoning to be successful, and for the paradox to arise, we must assume that nuclear retaliation is morally wrong. One might claim that nuclear retaliation is morally right, if it would destroy the opponents remaining missiles, or ensure the survival of the agent's nation. If this was the case, however, the non-consequentialist line of reasoning would no longer hold that forming the intention was morally wrong (since the act intended would be morally right) and the paradox would not arise.

I am interested in those cases where the paradox does arise. Therefore, rather than becoming entangled in the debate about whether forming the conditional intention to retaliate is the best way to prevent nuclear attack, or whether nuclear retaliation is always morally wrong, I have assumed that forming the conditional intention is the best way to prevent attack and that nuclear retaliation is morally wrong. By making these assumptions, I've been able to concentrate more directly on the paradox itself.

Quine has observed that,

"The argument that sustains a paradox may expose the absurdity of a buried premise or of some preconception previously reckoned as central to physical theory, to mathematics or to the thinking process.... More than once in history the discovery of paradox has been the occasion for major reconstruction at the foundations of thought."
With these comments in mind, my method in studying this paradox has been to examine the two lines of reasoning that sustain it. I suspected that the appearance of the paradox indicated that there was a premise or preconception in one, or both, of the lines of reasoning that should be rejected. If such a premise could be identified and addressed, then the paradox could be solved.

Allowing the assumptions that I've suggested above, I think that the central question that breeds the paradox is whether there can be circumstances in which it is morally right to intend to do an act that is morally wrong. The consequentialist line of reasoning holds that there can be such circumstances, while the non-consequentialist line of reasoning holds there cannot. Hence the paradox.

In Chapter 2, I concentrated on the consequentialist line of reasoning, which I found to be very plausible. This is not to claim that the consequentialist perspective is true, but rather that within commonly discussed consequentialist moral systems, it is consistent for the conditional intention to retaliate to be morally right and retaliation itself to be morally wrong. This line of reasoning treats the formation of the conditional intention to retaliate, and the act of retaliation, as separate acts, each with a unique set of consequences. Therefore, the normative evaluation of the conditional intention to retaliate is based on the consequences of the act of forming the intention, and is independent of the normative evaluation of retaliation.

If we assume that forming the conditional intention to retaliate is the best way to prevent a nuclear first strike, then it is consistent within the

---

consequentialist line of reasoning for the conditional intention to be morally right, even if the act intended, retaliation, is morally wrong. Whether or not forming the conditional intention to retaliate is, in fact, the best way to prevent nuclear war is a different question. My claim is that the first line of reasoning that leads to paradox is plausible and consistent.

In Chapter 3, I concentrated on the non-consequentialist line of reasoning. I think there are serious difficulties with this line of reasoning that are the cause of the paradox. This line of reasoning argues two points. First, that nuclear retaliation is morally wrong. As I've suggested, paradoxical deterrence requires this claim. Second, it is claimed that if retaliation is morally wrong then the conditional intention to retaliate must also be morally wrong. I've labeled the principle that supports this claim as the Wrongful Intentions Principle (WIP).

I've suggested two problems with this line of reasoning. First is that this line of reasoning does not address those cases where all the alternatives available to an agent violate some non-consequentialist principle. Arguments such as A1, that utilize the WIP, ignore this possibility and require rejection of the conditional intention to retaliate without consideration of the normative status of any of the alternatives.

A more important objection is directed at the WIP itself, which holds that the normative evaluation of the intention to do an act depends solely on the normative evaluation of the act itself. The WIP is required for the non-consequentialist line of reasoning to work, but I think there are good reasons to reject application of the WIP to the conditional intention to retaliate. It seems to me that paradoxical deterrence is a case which is not consistent with what we commonly take to be the relationship between an
action and the intention to do that action. Therefore, the assumption that the WIP can be applied to the case of nuclear deterrence seems to be the root of the paradox. Or in Quine's terminology, it is the preconception that must be challenged in order to explain the paradox.

In Chapter 3, I've shown how, in nuclear deterrence, the characteristics of the relationship between the conditional intention to retaliate and the act of retaliation are much different from what we normally believe the relationship to be between an action and the intention to do the action. In most cases, we believe that when an agent forms the intention to do an action, the agent desires the action to come about. The intention may also be seen as initiating a sequence of actions with the goal being the doing of the action intended. In such a case, the agent's purpose in forming the intention is to bring about the action. Finally, it is often the case that the only noticeable effect or consequence of forming the intention to do an act is the occurrence of the act itself.

For intentions with these characteristics, our intuitions seem to support the WIP, and thus, it seems plausible to hold that the intention is also morally wrong. We also may feel that the agent who forms an intention to do an act, with full knowledge that the only consequence of forming the intention is to bring about a morally wrong act, is demonstrating his desire to do the wrong action, and thus, has some moral defect in his character that drives his desire to bring about morally wrong actions. In such a circumstance, it seems plausible to claim that the formation of the intention is a morally wrong act.

These features that make the WIP intuitively appealing are not present in the case of the conditional intention to retaliate that is part of nuclear deterrence. The agent who forms the conditional intention to retaliate does
not desire to bring about the act of retaliation; he desires to prevent a nuclear attack against his nation. He believes that forming the conditional intention to retaliate is the best means of preventing such an attack. Thus, he believes that he is fulfilling a moral imperative to protect the citizens of his nation from nuclear attack by adopting the policy of nuclear deterrence; his purpose is to prevent nuclear attack. In such a case, the formation of the conditional intention may actually save innocent lives, even though the intended act would take many lives.

The fact that the intention to retaliate is a conditional intention plays an important part here. The agent's belief that the conditional intention to retaliate will prevent a nuclear attack means that he believes that the antecedent condition of the conditional intention will never obtain. Thus, he forms the intention to do an act that he knows would be wrong, with the belief that he will never have to do the wrong act; the formation of the conditional intention will ensure that the antecedent condition will not occur, or will have a very low probability of occurrence.

It also seems that these independent effects of forming the intention, and the fact that the intention is not formed for the purpose of bringing about the act, require that the intention and the action be treated as two separate actions. Also, the fact that, in the case of conditional intentions, the formation of the intention is usually separated temporally from the action intended, and the fact that if the antecedent condition does not occur, then the act will never be performed, also support the claim that the formation of the intention and the action intended are separate acts. If these are two separate actions, then it seems plausible to claim that a normative evaluation can be made of the act of forming the intention independently of the normative evaluation of retaliation.
The difficulty in the non-consequentialist line of reasoning is that the WIP does not provide an independent normative evaluation of the intention. Rather, it simply assigns the normative evaluation of the act intended, to the intention to do the act. Nuclear deterrence, however, is a case where the act of forming an intention may have the best consequences of all the available alternative actions (including not forming the intention), or may fulfill a non-consequentialist moral imperative. Yet, the WIP ignores these factors in the normative evaluation of the act of forming the intention to retaliate. Therefore, I want to reject the WIP as a premise in the non-consequentialist argument. Without the WIP, this line of reasoning breaks down and the paradox is resolved.

One response to this might be to attempt a reformulation of the WIP, that would take into account the unique features of the conditional intention to retaliate in the normative evaluation of forming the intention. If the WIP could be reformulated in this way, then the paradox might be revived. I don't see how this could be done.

I think that the WIP is the preconception about intentions and actions that creates the paradox. The WIP asserts that there is some type of special relationship between an action and the intention to do the action such that if the action is morally wrong then forming the intention must also be morally wrong. Paradoxical deterrence is a case where this supposed special relationship does not seem intuitively obvious, and in fact, seems counter-intuitive. We may have never given much thought to the truth of the WIP, until presented with a case like paradoxical deterrence, where it seems wrong to apply the WIP. Until the objector could specify exactly what the relationship is, and how it can account for the unique characteristics found
in the case of the conditional intention to retaliate, then the WIP needs to be rejected.

It might also be objected that regardless of the applicability of the WIP, a morally good agent could not sincerely form the intention to do a wrong action. I think that the independent effects of the conditional intention to retaliate are what make it possible for the morally good agent to sincerely intend retaliation, even when he knows that retaliation would be morally wrong. The agent believes that he will never have to act on the intention, because he believes that the formation of the intention will prevent nuclear attack (the antecedent condition of the conditional intention will not obtain). Likewise, the agent's purpose in adopting the intention does not include the act of retaliation. If his opponent attacks in spite of his deterrent policy, then retaliation does not contribute to his project-- his project of preventing attack, based on forming the conditional intention, has already failed. Retaliation could not contribute to the project of preventing nuclear attack once the attack has taken place.

It's also true that formation of the conditional intention is not an irreversible policy adopted by the agent. We can expect the morally good agent to be continually updating his assessment of the normative status of the actions available to him as the situation changes. Should an attack occur, it seems plausible to me to hold that the agent still has the option of not retaliating, even though he had previously formed the sincere intention to retaliate. I imagine that this is the case in the real life nuclear deterrence practiced by the US. The President, and certainly the missile launch officers in the ICBM silos, may have formed the sincere, conditional intention to
retaliate if the US was attacked with nuclear weapons. They really intend to retaliate.

But if an attack occurs, it's not necessary that they retaliate. Perhaps the nuclear destruction is even greater than they had imagined. Once they see the destruction, perhaps they will reassess their plan and decide that retaliation will serve no purpose. Or perhaps they will launch the retaliatory strike as intended. The point is that even though they have formed a sincere intention to retaliate, they are not locked into that policy as they would have been after, for example, activating an automatic retaliation machine that could not be turned off.

Thus, I think it can be plausible to imagine a morally good person who can form the sincere intention to do a morally wrong act. But I think the plausibility of this claim depends importantly on the independent effects of the conditional intention to retaliate. If these factors were not present, then it seems unlikely that the agent could sincerely form the intention.

Rejecting the WIP does not require rejecting the entire non-consequentialist line of reasoning. Non-consequentialist principles still might be applied to the conditional intention to retaliate. For example, if one of the principles is that the deliberate killing of innocent civilians is forbidden, then we need to ask whether the formation of the conditional intention to retaliate would result in the deliberate killing of innocent civilians. I don't think that in the forty years of nuclear deterrence, the conditional intention to retaliate has resulted in the deliberate killing of any innocents, thus forming the conditional intention does not seem to violate this principle.
If there were some principle which held that it is forbidden for government leaders to subject the citizens of their nation to excessive risk of nuclear attack, then the debate would be on how well nuclear deterrence can prevent attack. This would call into question our assumption that nuclear deterrence was the best means of preventing nuclear attack. But I think that this is where the debate should be—on how well the intention can prevent attack, or whether there is another policy that could more effectively prevent nuclear attack.

It seems misguided to me to reject forming the conditional intention to retaliate solely on the basis of the normative status of the act of retaliation, an act which may never occur, rather than assessing the conditional intention to retaliate on its own merits. Therefore, I think that nuclear deterrence is not "paradoxical deterrence" because we can resolve the paradox in paradoxical deterrence by rejecting the WIP and revising our preconceptions about the nature of the relationship between an action and the conditional intention to do the action.
BIBLIOGRAPHY


Lango, John W. "Is It Wrong To Intend To Do That Which It Is Wrong To Do?" *Monist*, 70, (July 1987) 316-29.


