UC Santa Cruz
UC Santa Cruz Electronic Theses and Dissertations

Title
Impact of Psychological Research on Moral Theory: Prospect Theory and the Doctrine of Doing and Allowing

Permalink
https://escholarship.org/uc/item/3n00j92k

Author
Dreisbach, Sandra Lyn

Publication Date
2012

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA

SANTA CRUZ

IMPACT OF PSYCHOLOGICAL RESEARCH ON MORAL THEORY

Prospect Theory and The Doctrine of Doing and Allowing

A dissertation submitted in partial satisfaction
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

PHILOSOPHY

by

Sandra L. Dreisbach

September 2012

The Dissertation of Sandra L. Dreisbach
is approved:

_____
Professor Daniel E. Guevara, Chair

_____
Professor Jorge Hankamer

_____
Professor Paul A. Roth

_____
Tyrus Miller
Vice Provost and Dean of Graduate Studies

Copyright © by

Sandra L. Dreisbach

2012

TABLE OF CONTENTS

1. INTRODUCTION 1

2. THE BASIC ELEMENTS OF TWO THEORIES: EXPECTED UTILITY THEORY AND PROSPECT THEORY 8

2.1. EXPECTED UTILITY THEORY 8
Gambling Scenario 8
2.2 PROSPECT THEORY 10
Richer Decision Problem 12
2.3 THE MECHANICS OF PROSPECT THEORY 15
2.4 PROSPECT THEORY'S EXPLANATION OF THE RICHER DECISION PROBLEM RESULTS 22
Summary of Group II Participant Decision Described by Prospect Theory 28

3. PROSPECT THEORY'S IMPLICATIONS FOR MORAL PHILOSOPHY 30

3.1 INTRODUCTION TO THE TROLLEY PROBLEM 31
The Trolley Problem 32
3.2 UTILITARIANISM 35
3.3 ISSUES FOR UTILITARIANISM AND CONSEQUENTIALISM RAISED BY THE INTUITIVE BASIS OF NON-CONSEQUENTIALIST THINKING 40
3.4 THE DOCTRINE OF DOING AND ALLOWING: A FUNDAMENTAL NON-CONSEQUENTIALIST ETHICAL PRINCIPLE 43
3.5 FOOT'S RESCUE DILEMMAS I AND II: DOING AND ALLOWING ILLUSTRATED 45
Foot's Rescue Dilemma I: 48
Foot's Rescue Dilemma II: 48
3.6 THE DOCTRINE OF DOUBLE EFFECT: ANOTHER FUNDAMENTAL NON-CONSEQUENTIALIST PRINCIPLE 50
3.7 FOOT'S TRADITIONAL ACCOUNT OF THE DOCTRINE OF DOING AND ALLOWING AND THE TROLLEY PROBLEM 53
3.8 QUINN ON DOCTRINE OF DOING AND ALLOWING AND DOCTRINE OF DOUBLE EFFECT 57
Neighbor's Furnace Case 62
Quinn's Rescue Dilemma III (a variant of Foot's Rescue II): 64
Quinn's Rescue IV (a variant of Foot's Rescue I): 66
3.9 QUINN'S FINAL RESOLUTION OF ORIGINAL TROLLEY PROBLEM 70

4. IMPLICATIONS OF PROSPECT THEORY FOR NONCONSEQUENTIALISTS 76

4.1 APPLICATION OF PROSPECT THEORY VS. THE DOCTRINE OF DOING AND ALLOWING 76
4.2 ASIAN DISEASE PROBLEM 77
Asian Disease Problem 81

4.3 QUINN'S RESCUE DILEMMAS AND HOROWITZ 82
Horowitz's breakdown of Foot's Rescue Dilemma I: 83
Horowitz's breakdown of Foot's Rescue Dilemma II: 83
4.4 ASIAN DISEASE DECISION PROBLEM CRITIQUE: EQUALIZATION 90
Group 1's Program Options 97
Group 1 Vaccine Interpretation of Asian Disease Problem 98
Group 2's Program Options 100
Group 2 Treatment Interpretation of Asian Disease Problem 102
4.5 ADDITIONAL EVIDENCE OF EQUALIZATION PROBLEMS 106
Snow Shovel Loss Case 108
Snow Shovel No-Gain Case 108
Categories of "Combat the Disease" Interpretations 110
"Combat the disease" Program Targets 112
Some "Combat the Disease" Means Interpretations 114

5. SUMMARY AND CONCLUDING REMARKS 120

APPENDIX: DILEMMAS AND CASES 125

APPENDIX: TERMS 133

BIBLIOGRAPHY 134

ABSTRACT

Impact of Psychological Research on Moral Theory:

Prospect Theory and the Doctrine of Doing and Allowing

by

SANDRA L. DREISBACH

It is widely acknowledged that Daniel Kahneman and Amos Tversky's Prospect Theory offers a better explanation of our decision making under uncertainty than some of the most widely accepted theories to-date, including especially Expected Utility Theory. I give a sympathetic summary of the case against Expected Utility Theory, but then show how Prospect Theory has not demonstrated that it can give a better explanation of our moral decision making under uncertainty (contrary, e.g., to the claims of Tamara Horowitz). In particular, it has not been demonstrated that it can give a better explanation of our intuitions in hard cases, such as the much discussed Trolley Car cases concocted by Philippa Foot and developed by her and Warren Quinn, among other contemporary moral philosophers concerned with issues often studied by both Expected Utility Theory and Prospect Theory. Prospect Theory has a weakness in common with Expected Utility Theory, namely, both are too focused on the consequences of our actions. In this respect, Prospect Theory is a consequentialist theory of moral decision making, and I bring out Prospect Theory's limitations by exploring the work of Foot's and Quinn's non-consequentialist theorizing in this context, with special emphasis on the Doctrine of Doing and Allowing, and related principles like the Doctrine of Double Effect.

DEDICATION

This dissertation is dedicated to my father, Lynn Wayne Garrison, who believed strongly in a good education and that no knowledge gained is useless. Though he claimed his father was the philosopher and not him, to me my father was a philosopher king.

ACKNOWLEDGMENTS

I must acknowledge several individuals whose efforts on my behalf are as much a part of this dissertation as my own efforts. Their generosity, hard work, faith and kindness helped to make this work possible. First and foremost, I must thank my family. They are my inspiration and my most fervent supporters. Thank you also to all of my fellow graduate students and friends who weathered the many ups and downs of this process and were a consistent voice of encouragement. I am also very much indebted to my main advisor, Dan Guevara, who tirelessly helped to guide and shape this work. The dissertation and I both are much the better for his dedication. A special thank you as well as to the rest of my dissertation committee, Paul Roth and Jorge Hankamer, as well as members of my qualifying exam committee: Ellen Suckiel, David Hoy, and Sandra Chung. Ellen Suckiel in particular has not just been an advisor but a true confidante and for that I am truly thankful. Thank you also to the Philosophy Department administrators who are always the unsung heroes. Last but never least I must thank my husband, Matthew Dreisbach, who has always believed in me, and my two sons, Sage and River Dreisbach, my companions for this journey.

1. Introduction

We like to think that we know the reasons for our choices. With moral decisions, in particular, where the stakes can be high and the ultimate consequences uncertain, we often take extra care in making sure that our reasons are good ones and that our decisions are justified. If, for instance, you are told there is an outbreak of a disease that is expected to kill 600 people and you need to decide between two competing disease combating programs, Program A and Program B, you will want to have some confidence that your decision is the best one, and that will involve making sure that it is well grounded. But imagine after choosing, let's say, disease combating Program A to combat the disease outbreak, you are told by psychologists who research such things that the only reason you picked Program A instead of Program B was due to seemingly insignificant changes in how the programs were presented to you. If Program A and Program B had been worded slightly differently, the pyschologists inform you, you would have chosen Program B instead. An arbitrary framing of Program A and B in one way instead of another, in other words, was the real basis for your decision. The Nobel Prize-winning work of Daniel Kahneman and Amos Tversky's Prospect Theory—a central focus of this dissertation—makes essentially this argument.[1]

Apparently, much decision making—while at first glance a seemingly straightforward process of making a well-considered choice between two or more

[1] Kahneman and Tversky 1979.

options—is neither straightforward nor obviously related to one's conscious deliberations. According to Prospect Theory, and other work in the social sciences, we come to see this somewhat disturbing fact when we attempt a scientific understanding of how and why we make the choices we make everyday. Theorists from diverse disciplines—ranging from Economics and Psychology to Mathematics and Philosophy—have each developed their own methods for understanding human decision making. Under the broad umbrella of decision theory, decision theorists attempt to not just describe our decsion making processes, but also focus on predicting how and explaining why rational agents make the decisions they do.

When we make a decision, on occasion we know with certainty what will be the result, more typically, however, we do not know with certainty the consequences of our choices and ensuing actions. For instance, if you are asked to close your eyes and draw a white marble from a clear vase that you can see in advance contains only one white marble, you can be rather certain that the marble you will pull out from the vase will in fact be a white marble. On the other hand, if instead you see two marbles, one white marble and one black marble, and you blindly reach into the vase, you may still pull out a white marble but you will not know this for certain beforehand. As everyone knows, you have a 50% chance of pulling out a white marble and a 50% chance of pulling out a black marble. And so many theorists over the centuries when dealing with decisions involving uncertainty, treat uncertainty as something to be measured by probabilities. As in our marble example, decision

theorists would describe the decision making as being a process of selection from among outcomes with different probabilities.

The probability of any particular outcome, however, is not enough to explain a decision maker's choice, it is also important to understand the reasons for the choice a decision maker makes. Let's go back to our marbles again, but this time we will stipulate that there are two containers. Container A has two white marbles and one black marble, and Container B has two black marbles and one white marble. If you want to pull out a black marble and can only choose one container to draw from, then, like most people, you will choose to draw from Container B since you have only a 1/3 chance to draw a black marble from Container A but a 2/3 chance to draw a black marble from Container B.

The reason you might give for your choice would perhaps be that you based your decision on what would most likely result in drawing a black marble. The decision theorist would put the same idea this way: you chose Container B because it would best maximize the *expected utility* of outcomes: utility being the desired outcome, which in this case is of course the drawing of a black marble. Accordingly, the traditional view of Expected Utility Theory was developed by decision theorists to not just describe certain decisions but also prescribe which choice ought to be made.[2] A rational decision maker in other words ought to choose based on what will best

[2] The traditional view of Expected Utility Theory is based on the work of J. Von Neumann and O. Morgenstern. See Morgenstern and Von Neumann 1947.

maximize expected utility.[3] If you had chosen Container A, which has fewer black marbles in it, instead of Container B, which has more black marbles, an expected utility theorist would argue that you did not make the optimal or best choice.

It is on the basis of this simple model that Expected Utility Theory became perhaps the most influencial theory of human decision making we have. But, as one may have already considered, many decisions are neither this simple nor straightforward and, so ultimately traditional Expected Utility Theory cannot sufficiently explain all human decisions. In reaction to shortcomings of the theory, there are have been several challenges to Expected Utility theory by decision theorists. The most important to-date is the theory I've already alluded to: Prospect Theory.[4] When psychologists Daniel Kahneman and Amos Tversky presented Prospect Theory, it caused a paradigm shift in decision theory and not just in the field of psychology but also in economics and across academic disciplines. As we will see, in chapter 2, sections 2.2, 2.3, and 2.4, the main innovation has to do with the fact that we seem to be more influenced in our decisions by (often) minor variations in how the decisions put to us are framed or phrased. How exactly the framing works to influence us is part of a rather complicated story that we will unravel later; it is a story that has become more complicated by how the various fields that study decision making in systematic and precise ways (again, economics, psychology, among others)

[3] Kahneman and Tversky 1979, p. 264.
[4] Prospect Theory aims at providing a better explanation of decision-making when dealing with risk (choices with known probabilities) than traditional Expected Utility Theory. See Horowitz 1998 where this is discussed at length, especially p. 369.

have tried to come to grips with it, especially in relation to counter challenges from Expected Utility Theory (a highly developed theory in its own right, of course).

Fortunately, things are simplified for us in present study because we will be concerned with the effect of Prospect Theory on a fairly well circumscribed area of philosophy, more particularly: moral philosophy. This will make our investigation manageable because philosophers have only just begun to grapple with Prospect Theory's implications. The Asian Disease problem roughly outlined earlier (which is a famous and important case used by Prospect Theory) helps to illustrate how moral decision making theory is being challenged by Prospect Theory. In fact this particular Asian Disease case, which I discuss at length (in chapter 4, sections 4.2, 4.3, and 4.4), has been used to support arguments in favor of Prospect Theory replacing traditional moral theoretical notions. The main question investigated in this dissertation is whether—as Prospect Theory maintains—it offers a better explanation of our moral decision making under uncertainty than some of the best moral theories to-date.

I argue that it does not, i.e. that Prospect Theory cannot fully explain our moral decision-making and that while it does have something to offer for moral decision making, it cannot fully demonstrate a better explanation of our moral intuitions than moral theory and reasoning of a traditional, well-developed sort. In partciular, the sort developed by Philippa Foot and Warren Quinn, among other contemporary moral philosophers concerned with issues often studied by both

Expected Utility Theory and Prospect Theory. One very important conclusion that I hope to establish is that, for all its differences with Expected Utility Theory, Prospect Theory has a weakness in common with it, namely, both are too focused on the consequences of our actions. In this respect, Prospect Theory is what moral philosophers call a *consequentialist* theory of moral decision making, and I will bring out Prospect Theory's limitations (which are, I think, the limitations of any consequentialist theory, including Expected Utility Theory), by exploring the work of Foot and Quinn, often regarded as the best non-consequentialist (or, as it is sometimes called, deontological) moral theorizing in this context.

To do this we will need some background to understand Expected Utility Theory and Prospect Theory in general, as well as particular philosophical theories and distinctions including the difference between consequentialist and nonconsequentialist lines of thinking. Throughout the course of this study the emphasis will be on ethical dilemmas, such as Foot's famous Trolley Problem, now very widely discussed by philosophers and social scientists alike (even brain scientists are getting into the mix with brain scans of subjects asked to work through the Trolley Problem and its seeminly infinite variations). Foot's famous Trolley Problem (presented and discussed in section 3.1, below) serves, like Prospect Theory's Asian Disease case, as a useful and central case for the study of the principles and intuitions involved in moral decision making.

To begin I will layout a basic understanding of the underlying Expected

Utility Theory to which Prospect Theory offers a response (see chapter 2, sections 2.1, 2.2, and 2.3). Then I will turn to an illustration of its shortcomings in light of the empirical evidence advanced by Prospect Theory. The discussion will then turn to an overview of Prospect Theory's alternative account. This will set us up for looking at Prospect Theory in more detail, with special focus on the framing effects it exploits.

Then we will be able to turn to some of the possible implications of Prospect Theory for Moral Philosophy. First, we will see some of the general implications the theory has for the ongoing debate between consequentialist and nonconsequentialist approaches to moral theory. The most influencial consequentialist theory in moral philosophy is Utilitarianism, so we will use it to represent the traditional consequentialist line of reasoning and briefly discuss its relation to its close cousin, Expected Utility Theory. Our primary focus, however, will be Prospect Theory's implications for nonconsequentialist moral theory.

2. The Basic Elements of Two Theories: Expected Utility Theory and Prospect Theory

2.1. Expected Utility Theory

Imagine a gambling scenario where you are given the following choice between two options:

Gambling Scenario

Option A: 50% chance to win $1,000, 50% chance to win nothing

Option B: $450 for sure [5]

According to Expected Utility Theory a rational decision maker will calculate the expected utility values for both Option A and Option B. The utility value is the desired outcome, in this case: having money (the more the better). The expected utility value of an option is based on both the utility value (measured here in dollars, of course) and the probability of the desired outcome occuring. To determine the expected utility value for Option A you multiply the monetary outcome of $1,000 times the probability of the outcome occuring, which here is 50%. This calculation will be 0.5 x $1,000 which results in the expected utility value of $500 for Option A. Option B, on the other hand, has the monetary outcome of $450 with 100% probability since it is a certainty. The expected utility value is calculated then as 1 x

[5] Example derived from Kahneman and Tversky's Israeli currency hypothetical choice problem. Kahneman and Tversky 1979, p. 264.

$450 which, of course, is $450. Expected utility theory states that a rational person will choose Option A, which has the higher expected utility value of $500 over Option B's lower expected utility value of $450. For Expected Utility Theory, even though it is not certain that you will gain $1,000 by choosing Option A because it is a gamble, the expected utility value of $500 for Option A is $50 more than the expected utility value of $450 for Option B, and as such Option A maximizes the expected good consequences.[6] In general, then, the traditional view of Expected Utility Theory is that, if rational (and informed), we will always choose the action that is expected to maximize desired outcomes (i.e.utility) for ourselves and in this case that is accomplished by selecting Option A.

Yet in real decision-making situations, individuals frequently do not choose the outcome with the greater expected utility (even when informed). Staying with the same example, Kahneman and Tversky found that when people were presented with situations, such as the Gambling Scenario, they on the whole tended to prefer certain outcomes over uncertain risky outcomes, even when expected utility was greater for the latter. For example, they found that subjects tend to select Option B's certain $450 gain over Option A.[7] Despite Expected Utility Theory indicating that Option A

---

[6] Expected utility theory could make an argument for Option B if the scenario is a one time gamble versus a repeated gamble over time. However, Expect utility theory's reasons for this would not be based on Prospect Theory's view on risk. For more on this see Kahneman and Tversky 1979.

[7] Kahneman and Tversky report this finding in their original research that this particular scenario was based on: the Israeli currency hypothetical choice problem. See Kahneman and Tversky 1979, p. 264. This finding is not isolated and has been

is the choice that would best maximize expected utility, people have been found to refuse to make this sort of gamble. This decision, and decisions like it that do not follow the Expected Utility Theory model, lead decision theorists to puzzle about what basis people are making their decisions on if they are not based on expected utility. Some theorists have speculated that people simply are, generally, not rational.

2.2 Prospect Theory

Prospect Theory, on the other hand, puts forth the argument that actual decisions under risk are determined not purely by the expected utility of the choices alone, but instead also by a complex set of factors including an aversion to risk when considering gains (as opposed to a tendency to take risks to avoid losses), and by how people frame their possible choices relative to a neutral reference outcome which makes the choices appear as either a gain or a loss relative to the neutral reference outcome. None of these things, which I will explain eventually, are considered relevant by Expected Utility Theory. But the effect of considering them in decision making has had a huge effect. In the period of 1975-2000, Kahneman and Tversky's 1979 "Prospect Theory: An analysis of decision under risk"[8] was the second most cited paper in economics and in 2004 the work earned Kahneman the Nobel Prize in

---

confirmed numerous times including by more recent studies such as "Problem 3" and "Problem 4" gambling scenarios in Kahneman and Tversky 1981, p. 454-455.

[8] Kahneman and Tversky 1979.

Economics.[9] Prospect Theory has not just caused paradigm shifts in both the fields of Psychology and Economics but its influence has also been far-reaching, impacting other fields of study. Psychologists Reid Hastie and Robyn Dawes even indicated in 2001 that "Prospect theory is the best comprehensive description we can give of the decision process. It summarizes several centuries' worth of findings and insights concerning human decision behavior. Moreover, it has produced an unmatched yield of new insights and predictions of human behavior in decision making."[10]

Even though Kahneman and Tversky's work on Prospect Theory is essentially a descriptive account of decion-making under risk, there are also significant implications for philosophy and in particular for moral theory. Kahneman and Tversky themselves even pointed out in their 1998 work "The Framing of Decisions and the Psychology of Choice" that "the psychology of choice is also relevant to the normative question of how decisions ought to be made" and that "when framing influences the experience of consequences, the adoption of a decision frame is an ethically significant act."[11] Moral philosophers have just begun to do the necessary work to understand the implications of Prospect Theory for ethical theory. Eventually, I will investigate the specific question of whether Prospect Theory provides a better explanation than certain moral theories in accounting for moral decision-making under risk. But for now let us continue to lay out Prospect Theory by

[9] Gonzalez, Wu and Zhang 2004.
[10] Dawes and Hastie 2001, p. 310.
[11] Kahneman and Tversky 1998, p. 458.

turning to another case that will help us illustrate in more detail how the theory works.

We turn, then, to a well known case from Kahneman and Tversky's original research on Prospect Theory,[12] a case I will call the "Richer Decision" problem. In the experiment, there are two separate groups of subjects, referred to here as Group I and Group II, that are each presented the same problem but with slightly different wording. I will start here with Group II, for reasons that will become clear later. Group II participants are told to imagine that they are given $500 and then presented with two options. Participants can choose Option One, where they will certainly lose $100, – or – they can choose Option Two, where they will have a 50% chance of losing nothing but a 50% chance of losing $200.

Richer Decision Problem

Group II
Setup: Participants told they are given $500

Option One: Lose $100

Option Two: 50% Lose Nothing, 50% Lose $200[13]

Kahneman and Tversky found that most people in Group II tended to choose Option Two where, even though you may loose $200, you still have a 50% chance of

[12] Tamara Horowitz (one of the chief exponents of Prospect Theory) provides this example in Horowitz 1998, p. 369-70.
[13] A complete list of cases can be found in the appendix for ease of reference.

loosing nothing. Option Two seems better to most of the participants in Group II, who were being evaluated, than the certain loss of $100 in Option One. Expected Utility Theory, however, would have predicted that the participants, if rational, would have preferred Option One instead. Since, if you look at just the expected dollar amounts, with Option One you stand to have $400 for certain but with Option Two you actually risk ending up with only $300. If you are making your choice only on what would most likely maximize your money, you ought to choose Option One for the certain $400 and not risk loosing an additional $100 by taking the Option Two gamble of the 50% shot at $500. Prospect Theory though can explain why participants in Group II tended to pick Option Two instead despite the choice being contrary to what Expected Utility Theory would predict for informed and rational participants.

Before we go on to explain Prospect Theory and provide Prospect Theory's explanation of Group II's decision, it is important to contrast Group II with Group I in the Richer Decision Problem. Group I is told instead to assume they have been given $300 and then presented with the choice of either a certain gain of $100 (Option One) or a 50% chance of either gaining nothing or a 50% chance of gaining $200 (Option Two). Group I tended to pick their version of Option One rather than their version of Option Two, in contrast to Group II.

Richer Decision Problem

Group I
Setup: Participants told they are given $300

Option One: Gain $100

Option Two: 50% Gain Nothing, 50% Gain $200

This is significant because in the Richer Decision Problem the subjects in Group I and Group II are both presented with two choices that have the exact same expected utility values: namely the certain outcome that they will have $400 if they choose their respective Option One's, or the uncertain, but equivalent, expected outcomes of a 50% chance of either having $300 or $500 in Option Two. If Expected Utility Theory were correct, regardless of which group the participants are in, they should all tend to choose their Option One because the two choices for both Group I and Group II are mathematically equivalent in terms of expected monetary outcomes. So if their decision is being made by what they determined would maximize money, and perhaps by quickly calculating the numbers of each option and examining the probabilities before choosing, the participants of both groups would know that Option One's outcome of a certain $400 is the best choice in terms of expected utility, and everyone should choose the first option. Despite this equivalency, the research

demonstrates that the two options are not given equal consideration by the two groups of individuals. [14]

2.3 The Mechanics of Prospect Theory

Prospect Theory tries to give a reasonable explanation of why these subjects chose as they did. Prospect Theory maintains that, instead of strictly evaluating which option has the best expected utility, subjects first frame their possible choices based on a reference point and then from that reference point decide which option has the best expected utility. In other words, there are two phases of the choice process according to Prospect Theory.

The first phase is the editing phase where a person takes the options given to them, conducts an initial analysis of them and then rearticulates them in a reduced and easier form. As Kahneman and Tversky say, "the function of the editing phase is to organize and reformulate the options so as to simplify subsequent evaluation and choice. Editing consists of the application of several operations that transform the outcomes and probabilities associated with the offered prospects."[15] While the editing

[14] Kahneman and Tversky 1979.

[15] Kahneman and Tversky 1979, p. 274. There are four major operations of the editing phase including: coding, combination, segregation, and cancellation. In coding each prospect is evaluated as either being a gain or a loss relative to the neutral reference outcome. The combination operation prospects can "sometimes be simplified by combining the probabilities associated with identical outcomes." Segregation is where a prospect's riskless component is segregated from the risky aspect. Cancellation is

phase breaks down into several components, it is important to focus on primarily the framing operation[16], which may prove the most relevant for evaluation of moral decision-making.

Framing is the part of the editing phase that determines which of the various outcomes are seen as the neutral reference outcome.[17] The agent examines each of the possible outcomes and selects one as the neutral reference outcome. This neutral reference outcome then is used as the basis from which to evaluate all of the various choices. Tamara Horowitz—one of the most important expositors of the theory, especially in application to moral philosophy (see chapter 4, section 4.3 below) notes that it is apparently "an open question as to what factors determine an agent's selection of a neutral outcome, although it often seems to correspond to the status quo."[18] I will assume with Horowitz for now that what "is most important for our purposes, though, is that different formulations of a decision problem can lead an

---

where aspects that are shared by different prospects are discarded which results in what Kahneman and Tversky describe as isolation effects. There are also two additional operations of "simplification and the detection of dominance. The first refers to the simplification of prospects by rounding probabilities or outcomes. … A particularly important form of simplification involves the discarding of extremely unlikely outcomes. The second operation involves the scanning of offered prospects to detect dominated alternatives, which are rejected without further evaluation." Kahneman and Tversky 1979, p. 275.

[16] I am following the similar move made by Horowitz to focus on the framing aspect. For more on the details of each operation of the editing phase, see Kahneman and Tversky 1979, p. 274-5.

[17] Horowitz 1998, p. 371.

[18] Horowitz 1998, p. 371. Note that Horowitz here refers to "neutral reference outcome" as simply the "neutral outcome."

agent to make different choices of the neutral outcome."[19] Prospect Theory holds that after the establishing of the reference point, choices will be categorized as a gain or a loss on the basis of whether they would be more beneficial (gain) relative to the neutral reference outcome or detrimental (loss). So to clarify how framing edits the options in Prospect Theory, let's go back to the Richer Decision problem.

In the experiment of the "Richer Decision" problem, the participants in Group I are told they are given $300 before they make their choice. This $300 becomes the neutral reference outcome which is assigned the baseline value of zero. This is due to the subjects viewing this newly acquired $300 as "theirs" and becomes, from their point of view, a part of whatever money that they now have before they make their choice. In other words, according to the theory this becomes, in effect, the status quo (although I will question this sort of determination of the status quo later). This $300 then is used as a neutral reference outcome from which to edit the options as being either a gain or a loss. If a particular choice has an outcome that results in an increase of money that is more than $300 then it is seen as a "gain." But, if a particular choice has the outcome of decreasing the total amount of money under $300, then it is viewed as a "loss." Any option under $300 will be considered a "loss" even if the participant would have more money than what they had before the $300 starting money. For instance, if they end up with just $200 this would still be seen as a "loss" because it is less than the given $300. According to Prospect Theory, participants in

[19] Horowitz 1998, p. 371.

Group I must have therefore edited their options into a choice between (i) a gain of $100 for Option One relative to the $300 neutral reference outcome and (ii) a 50% chance of a loss of $100 and 50% chance of a gain of $200, given the $300 neutral reference outcome.

The edited choices, then, are not based on whether or not they will have $300 in the end, but on how much more they can gain or lose from the established $300 neutral reference outcome. To see how this setting of the neutral reference outcome can vary, recall the contrast between Group I and Group II in the Richer Decision problem example. While the first group was given $300 to begin with, the second group was given $500. Prospect Theory maintains that this difference in starting money creates a difference in what Group II takes as the neutral reference outcome. Individuals in Group II view $500 as their neutral reference outcome instead of what Group I establishes as their neutral reference outcome, $300. Participants, therefore, in Group II had to edit their options into a choice between (i) a loss of $100 for Option One relative to the $500 starting point and (ii) a 50% chance of keeping the $500 (losing nothing) and a 50% chance of a $200 loss, given the $500 neutral reference outcome. Note here that both Option One and Option Two will actually result in each participant having more money than what they started with if examined just in terms of expected utility since Option One's cash outcome is $400 and Option Two's possible outcome will be: 50% of receiving $500 or 50% of having $300. But this (i.e. judging according to expected utility) is exactly, according to Prospect

Theory, what participants do *not* do, since the outcomes are edited in the framing process of the editing phase as a gain or a loss relative to the neutral reference outcome and not examined strictly in terms of expected utility.

Once the editing phase clarifies the choices, also called "prospects," [20] the next phase of the process, the evaluation phase, rates each of the prospects in terms of desirability. In the evaluation phase, Horowitz says, the

> …agent encodes her judgments of desirability differences with a value function, *v*. The function *v* takes the value 0 for the neutral outcome and takes positive or negative real values for other outcomes in such a way as to reflect the positive or negative deviations in desirability of those outcomes from the neutral outcome. . … In deciding which actions are preferable to which other actions, agents do not simply multiply the value of outcomes by the probability of those outcomes. Some probabilities are factored in at more than their face value, while other probabilities are factored in at less than their face value. Or so a large body of experimental research seems to suggest. To reflect this fact, prospect theory attributes to the agent a weighting funtion, $w(p)$, which associates with each probability a "decision weight."[21]

While the details of how precisely to calculate the probabilities with their associated decision weight is important for Prospect Theoretic analysis[22], what is important to

---

[20] Prospects are typically possible actions that involve risk or gambles. Risk is used to refer to choices that are uncertainties with a probability for yielding particular outcome(s). A riskless prospect is a certainty. For more on this see Kahneman and Tversky 1979, p. 263.

[21] Horowitz 1998, p. 371.

[22] Horowitz 1998, p. 372, "Positive and negative prospects are evaluated in the following way. In the editing phase, they are segregated into two components, a riskless component which is the minimum loss or gain certain to be received and a risky component comprised of the additional gain or loss at stake. These prospects are then evaluated in the following way: if $p+q = 1$ and either $x > y > 0$, then $D(x,p; y, q) = v(y) + w(p)[v(x) - v(y)]$. This value equals the value of the riskless component, plus the difference between the two outcomes multiplied by the weight

understand for our purposes here is that, according to the theory, for each choice, if it is edited as a "loss" relative to the neutral reference outcome, the choice will be given greater weight in terms of its ranked value in the evaluation phase than if it is edited as a "gain." A "positive deviation" designates how much in the perceived positive direction or "gain" an outcome is from the neutral reference outcome. A "negative deviation" designates to what degree in the perceived negative direction or "loss" the results of a prospect would be from the neutral reference outcome.[23] Basically, Prospect Theory argues, according to Horowitz, that when evaluating their edited choices "people tend to be risk averse when it comes to [obtaining] gains, risk seeking when it comes to [avoiding] losses, and their response to losses tends to be more extreme than their response to gains."[24] Risk again means that the choices involve uncertainties so there is a probability assigned to each possible outcome versus a choice, or prospect, that involves a certain outcome. At the end of the evaluation phase, each edited choice has been given a rank in order of preference and the agent can then make a decision based on the choice that has the higher preference being selected over a lower ranking choice.

---

associated with the outcome with the greater absolute value." For more on this see Kahneman and Tversky, p. 276.

[23] Again, for more detail see Kahneman and Tversky 1979, p. 276.

[24] Horowitz 1998, p. 372. This may seem counterintuitive at first glance but in general this means that people are inclined to risk to avoid losses and are not inclined to risk and in fact are risk avoiding when it comes to garnering gains. For more on this see Kahneman and Tversky 1979, p. 279.

To show how the evaluation phase works in practice let's again go back to the Richer Decision problem. Once the participants in Group I have edited their choices as being between: (i) a gain of $100 for Option One relative to the $300 neutral reference outcome and (ii) a 50% chance of a loss of $100 and 50% chance of a gain of $200, given the $300 neutral reference outcome; they can clearly evaluate the two options. According to Prospect Theory, an agent ranks the certain $100 gain as being the preferable choice since it has been established as both a gain in the editing phase and as a certainty in terms of its decision weight.

Option Two, when compared to Option One, is not only uncertain in terms of its decision weight and, therefore, a risky prospect, but also the outcomes for each possible result in Option Two are not greater than the gain represented by the first option. Option Two is equally likely to either result in a loss of $100 or a gain of $200. The possible loss of $100 in Option Two is clearly not greater, strictly in terms of monetary gain, than Option One's certain gain of $100. The other possibility for Option Two is greater in terms of monetary gain, since the 50% chance of a $200 gain is greater than Option One's $100 gain. However, it is not greater in terms of decision weight since a certain $100 gain is (according to the theory) valued as more preferable to an 50% chance of a $200 gain (which also has an expected utility value of $100 since 0.5 x $200 is $100) because of a subject bias towards certainty (there is a larger decision weight on certain prospects than uncertain prospects). This is true even if the 50% chance of a $200 gain was not paired with the possibility of a $100

loss. In this case though the possible $200 gain in Option Two is paired with the possibility of a loss, and this combined risk only makes the decision weight value for the prospect of Option Two even worse when compared to Option One's certain $100 gain. In other words, participants in Group I were risk averse to the potential gain of $200 (or loss of $100) in Option Two and so chose the certain gain of $100 in Option One. This reflects Prospect Theory's notion that people will take greater risks to avoid losses than to garner gains.

2.4 Prospect Theory's Explanation of the Richer Decision Problem Results

With a general understanding of how both the evaluation phase and editing phase work in Prospect Theory to explain decision making under risk, we can turn now to how Prospect Theory can explain the decision made by participants in Group II in the Richer Decision problem and others like it. Recall that Expected Utility Theory would, as does Prospect Theory, predict that participants in Group I would select Option One.

However, Expected Utility Theory's reason for this differs from Prospect Theory's basically because the former explains things in terms of what rational and informed participants will tend to choose, namely, what would best maximize their expected monetary gain. Given this, Option One is clearly preferable for Group I. Option One offers a certain monetary gain of $100 versus risking a 50% chance of

monetary loss in Option Two.[25] Expected Utility Theory has no need to edit the options or evaluate them in terms of loss or gain relative to a neutral reference outcome, in contrast to Prospect Theory, and simply looks strictly to expected utility value of each choice, compares them, and selects Option One. Since the expected monetary outcomes of each choice do not change for Group II, Expected Utility Theory predicts again that participants in Group II should also pick their Option One. But this is where Expected Utility Theory falls short in describing decision making under risk. As Kahneman and Tversky's Richer Decision problem study shows, participants in Group II did not tend to pick Option One and instead tended to pick the less optimum, in terms of expected monetary values, Option Two.

There are two main reasons for Group II's participants' tendency to select Option Two instead of Option One, according to Prospect Theory. First, as indicated earlier, Group II set a different neutral reference outcome than Group I in the editing phase due to the different instructions given. The instructions were different both in terms of starting money and in how the two options were phrased. Group I was told they had $300 at the onset of the problem and therefore set their neutral reference outcome at $300. Whereas Group II set their neutral reference outcome at $500 based on their instructions.

[25] Since the expected utility values are equal, diminishing marginal value of wealth is what explains the preference for the sure-thing in Expected Utility. See Kahneman 2011, p. 274.

It is also worth noting that both of the Options' choices are framed for Group II as a "loss" versus the more positive phrasing as a "gain" which was done for Group I. The location of the neutral reference outcome can also be influenced by this sort of phrasing of the options as well, according to Prospect Theory. This is because of what Kahneman and Tversky call the isolation effect, which is the idea that "people generally discard components that are shared by all prospects under consideration" which "leads to inconsistent preferences when the same choice is presented in different forms."[26] This seems to occur in the editing phase of the decision process according to Kahneman and Tversky. In the Richer Decision case, the neutral reference outcome seems to be primarily affected by the starting money, but in a case we will look at later—the Asian Disease case (which was briefly discussed in the introduction, page 1 above)—different phrasing of the options themselves will play a larger role in the setting of the neutral reference outcome.

The significance of having two different neutral reference outcomes for Prospect Theory is that when the choices are evaluated they are given a preference order based on the neutral reference outcome set up in the editing phase. Since Group I and Group II set different neutral reference outcomes, this in turn will be reflected in the evaluation phase and, ultimately, having different neutral reference outcomes will be part of the reason for Group I and Group II participants tending to make different decisions.

---

[26] Kahneman and Tversky 1979, p. 263.

The other reason for Group II selecting Option Two, according to Prospect Theory, is due to different valuations of Option One and Option Two in the evaluation phase. The differing neutral reference outcomes of $300 and $500 set in the editing phase resulted in different valuation rankings for Option One and Option Two in the evaluation phase by each group. As discussed earlier for Group I, Option One's certain increase of $100 and Option Two's possible increase of $200 was determined to be a positive deviation from their neutral reference outcome of $300 in the editing phase and was therefore seen as a "gain." However, Option Two, while valued as a positive deviation from the neutral reference outcome, was not given greater decision weight in the evaluation phase in part because it is an uncertainty: there is a 50% chance of gaining $200 –or- a 50% chance of gaining nothing. The idea, according to Prospect Theory, is that we will avoid negative deviations before trying to secure positive deviations from the neutral reference outcome. Therefore, when evaluating Option One and Option Two, Option One is given a higher preference ranking over Option Two.

This is in contrast to how the two options are evaluated in Group II. For Group II, instead of both options being seen as "gains" as they were for Group I, both of the choices were seen as negative deviations from their neutral reference outcome of $500 in the editing phase resulting in each option being viewed as a "loss." Then, when Group II participants evaluated Option One and Option Two, Option Two is seen by them as more desirable than Option One because it offers the possibility of a

lesser "loss," the 50% chance of losing nothing. Group II leaned towards the riskier (in terms of expected utility) Option Two of a 50% chance of losing $200 –or- a 50% chance of losing nothing, because given their neutral reference outcome of $500 it is worth the risk. This reflects Prospect Theory's concept that individuals take greater risks to avoid losses than to garner gains.

So while participants in Group I were risk averse to Option Two, since it was evaluated as a risky "gain," participants in Group II were risk seeking by selecting Option Two, because both Option One and Option Two were evaluated as losses. These differences, for Prospect Theory, explain why a subject would make the risky choice of selecting Option Two in Group II, but not make the same choice in the other group, Group I, despite being given a mathematically equivalent set of ultimate outcomes. Clearly, something other than just expected utility is affecting what alternative the subjects select. The main difference in the experiment is in how the instructions were presented to each group and this difference is enough to change what a subject thought was the best choice. Decisions (especially under risk) seem to be determined in part by how people frame their possible choices relative to a neutral reference outcome based upon outcomes before them and not based purely on expected utility alone.[27] Apparently, decisions involving risk only need to be presented (framed) in the desired way, positively or negatively, for us to describe fairly accurately how people may actually choose, given that the framing of the

[27] There is a great deal of experimental evidence for this, see Kahneman and Tversky 1979.

choices determines the neutral reference outcome. Hence, the initial implications of Prospect Theory are not just that it is a more preferable alternative theory to Expected Utility Theory but that the way a subject frames an outcome affects the utility he or she expects.

Summary of Group II Participant Decision Described by Prospect Theory

Possible neutral reference outcome for Option Two choosing participants:

Acquiring five hundred dollars (+$500) at the start is taken up as part of the participant's status quo. It is not seen as an outcome but a part of their baseline.

Neutral reference outcome: +$500

Option One Framed:
A certain loss of $100. This is a reduction of $100 from the $500 neutral reference outcome, in other words, ending up with $400 is seen as a loss.

Option Two Framed:
A 50% chance of losing $200; and 50% losing nothing. Results in losing $200 making the end result of $300 a perceived loss relative to the $500 neutral reference outcome –or- maintaining the +$500 neutral reference outcome, appearing to be positive and more desirable when compared to the certain "loss" in Option One.

In being able to explain decisions like these, Prospect Theory has been a very persuasive alternative explanation to expected utility theory.[28] However, the influence of Prospect Theory does not end with decision theory in economics and psychology but it has also affected philosophy. There is a huge body of literature in the social sciences debating its merits there, but we will be focused on its philosophical merits.

[28] Hansson 2011.

3. Prospect Theory's Implications for Moral Philosophy

Prospect Theory has some significant implications for moral philosophy in particular.[29] Prospect Theory attempts to describe decision-making under risk and moral decision-making involves risk. Moral decisions are made where there are various uncertainties and often high stakes such as the loss of human life. So even just upon general reflection, it seems natural to think that Prospect Theory may have something of interest to moral philosophers.

Normative ethical theory in particular is concerned about what moral choices we ought to make especially when the situation is uncertain. Prospect theory's potential implications here call for especially close examination. In general, there are various moral philosophical approaches as to how to ascertain which choice among options at hand for a decision maker is the best or morally preferable one. For our purposes two historically significant moral perspectives will be examined in relation to Prospect Theory: consequentialist and non-consequentialist normative views. Each view will be discussed, and non-consequentialist perspectives at length, but for now it is important to show how decision making under risk is generally approached by normative moral theory. An example often used to illustrate these contrasting perspectives in moral philosophy and the issues they struggle with is the famous Trolley Problem.

---

[29] Other fields of philosophy should also make note of Prospect Theory but there seem to be larger implications for moral theory.

3.1 Introduction to the Trolley Problem

Here is the Trolley Problem: Imagine you are the driver of a trolley that is traveling out of control towards three innocent persons tied to the track.[30] There is not time to remove these people from the imminent danger of being killed, but you can steer the trolley away from them by pulling a lever and moving it onto an alternate track. However, this other track also has an innocent person tied to it who would be killed should the trolley be redirected. What would you do? Steer to the other track resulting in the death of one person? Or choose for the trolley to remain on its course, killing three people instead?

[30] Philippa Foot concocted the original Trolley Problem and I present essentially Foot's example, though not exactly in every detail. See Foot 2002a, p.23. There is a mountain of literature on this problem, only growing through a new wave of interest in it due to Prospect Theory's interest in it and related issues, and due also to other work coming of out the social sciences, and even now, the science of the brain. For a good survey with references see Howard-Snyder 2011.

The Trolley Problem

Imagine you are the driver of a trolley that is traveling out of control towards three innocent persons tied to the track. There is not time to remove these people from the imminent danger of being killed, but you can steer the trolley away from them by pulling a lever and moving it onto an alternate track. However, this other track also has an innocent person tied to it who would be killed should the trolley be redirected.

Both philosophers and non-philosophers alike are inclined to say it is better to save the three by switching the tracks instead of refraining from killing the one person.[31] But what is the moral justification for this decision? Perhaps we think if we truly cannot save all four people then we *ought* to save as many people as we reasonably can. Human life is valued for its own sake and we recognize the net value gained in lives by switching the tracks. We may even blame the person unwilling to switch the tracks for not acting properly. This would be the consequentialist line. But what will concern us is a non-consequentialist or deontological line of reasoning, according to which it is difficult to justify the killing of an innocent person simply in terms of number of innocent lives saved. In particular, we will be concerned with the leading non-consequentialist accounts: namely those of Philippa Foot and especially

[31] There is ongoing research but Joshua Greene, who has one of the most cited and influential studies, and others have found consistent responses on the trolley problem. See, e.g., Greene and Haidt 2002, p. 519. Cf. Lanteri, Chelini and Rizzello 2008.

Warren Quinn's development of Foot's account. These have been directly discussed by Prospect Theorists.

In general, non-consequentialists believe that there are certain moral constraints on the maximization of good consequences. The end of maximally good consequences does not necessarily justify the means when direct harm or killing of innocents is involved, for example.

This, again, is in contrast to consequentialist moral theories, which entirely determine the moral correctness of an act on the basis of the consequences of the act. The Trolley Problem is a difficult and important problem for non-consequentialists who, like Quinn, do not want to base their decision primarily on the consequences. A consequentialist, however, can more readily say, just as they will be able to do in many related tough cases (e.g. Foot's Rescue Dilemma's I and II to be discussed, below), the Trolley Problem is for the most part a non-issue for them since maximizing the consequence of more lives saved is already a consequentialist based moral decision. If Prospect Theory shows that we don't really rely on the non-consequentialist principles, such as those employed by Quinn and Foot, then this might provide confirmation of the consequentialist position against that of Quinn and others who wish to provide a principled non-consequentialist solution to the Trolley Problem.

Before turning to one such crucial principle, let me note that some may respond to the Trolley Problem by rejecting aspects of the thought experiment. One

might think that if we knew more about the people involved or the situation we might choose differently.[32] In this situation, however, there are no additional details and we are left to act with limited information and knowledge. It is worth noting that we must almost always perform moral actions with limited information and knowledge upon which to base our decisions. In addition, our hands have been tied in terms of possible action, being given only the choice to steer or not to steer the trolley. We would likely have more choices at our disposal in a *real* trolley situation.[33] But this is irrelevant to the Trolley Case we are trying to consider.

Yet even taking account of these concerns about using such thought experiments[34], it still appears plausible to say that we ought in this case to save as

---

[32] For instance, say the one person was the only one who knew the cure for cancer and could perhaps save millions of lives with this knowledge. Then we may instead choose for the three to die as this information about the one person gives us a reason to prefer saving him despite the fact that one would be choosing one person to live instead of three.

[33] By this I mean an experience one could have in reality with real trolleys and people. While the trolley problem is generally plausible as a thought experiment, it is an unlikely scenario at best that one would be in this particular situation, where innocent people would be tied to or perilously beside train tracks, and where one's options were so limited that the steering of the trolley onto an alternate track is the only free choice available. Also, as even Foot noted when she originally posed the problem, we do not have the same sense of certainty in real life scenarios as we do in this case. This is because we generally do not know many facts with absolute certainty such as that the person on the tracks would in fact be killed by the trolley.

[34] It has also been argued that thought experiments do not actually provide any real insight into moral decision-making as they are so divorced from everyday experience and moral practice. Thought experiments, however, can be useful in thinking about moral issues by allowing us to simplify known moral situations as well as imagine possible moral dilemmas. Simplification helps bring to light specific features of our decision-making that may be difficult to deduce from actual dilemmas. Imaginative thought experiments permit moral agents to think through possible actions before

many lives as we can by our actions, and our initial reaction to the Trolley Problem seems to illustrate this tendency in our moral decision-making. This initial take on the reasons for why we would choose to steer the trolley towards the one seems, as we have indicated, to give a consequentialist justification. The most cited consequentialist justification for redirecting the trolley has historically been the Utilitarian one. There are other consequentialist moral theories but for our purposes Utilitarianism will serve to illustrate the main consequentialist line of thinking in normative ethics.

3.2 Utilitarianism

Utilitarianism has been one of the most influential normative theories in ethics. Classical Utilitarianism was originally formulated by Jeremy Bentham and later refined by John Stuart Mill. The basic premise of Utilitarianism is that the moral correctness of an action is determined by whether the action maximizes good consequences.[35] So, for a Utilitarian, the reason saving more lives in the Trolley Problem is the morally correct action versus choosing to only save the one is because, all other things equal, saving more lives tends to lead to more good consequences overall than just saving one. Since the consequences of the action are the basis for

---

encountering them and in hopes of making better choices or simply avoiding types of situations altogether.

[35] What counts as a good consequence can vary depending on the consequentialist theory being considered. For Utilitarians a good consequence is based on utility which for traditional Utilitarianism was happiness, pleasure or satisfaction.

deciding whether the action is morally right or wrong, Utilitarianism, as mentioned, is considered to be a consequentialist theory. I will discuss consequentialist theories again in more detail later. For now what is important to understand about Consequentialism is that it is a category of moral theories which base moral correctness primarily through the examination of consequences.

Utilitarianism as a particular consequentialist theory also wishes to maximize the good consequences for all people.[36] This Utilitarian idea was famously coined by Jeremy Bentham as "the greatest happiness of the greatest number that is the measure of right and wrong."[37] Utilitarianism hence considers the interests not just of the individual self but of all people when evaluating the consequences of an action. Utilitarianism in this way is impartial and does not consider any one person's happiness as above anyone else's. Furthermore, Utilitarianism is agent-neutral such that whatever reason I may have to "promote the overall good is the same reason anyone else has to so promote the good. It is not particular to me."[38] So whatever reason I have to morally do or not do some particular action on a Utilitarian account must also be a reason for anyone else considering the same action.[39]

---

[36] This is also phrased as maximizing utility which will be discussed in greater detail later. In addition, I mention people here but Utilitarianism can consider nonhuman animals as well.
[37] Bentham 1776.
[38] Driver 2009.
[39] It is worth noting here that there are two main types of Utilitarianism: Act Utilitarianism and Rule Utilitarianism. When Utilitarianism is generally discussed, Act Utilitarianism is usually what is being referred to. Act Utilitarianism determines the greater good by examining each individual act's consequences whereas Rule

A Utilitarian could try to argue, for instance, that we should not switch the tracks towards the one because the one person's death may have more negative consequences than the deaths of the three innocent persons trapped on the track. This would not be sufficient reason on Utilitarian grounds because we do not know in the limited scenario provided by the Trolley Problem whether the one person's death would lead to more negative consequences but we do know that in general killing more people than one has to is a clearly greater negative consequence than not. For a utilitarian, there needs to be a greater good that would be served from not switching the tracks and any good that results must outweigh the negative consequences from the act. Choosing to not switch the tracks towards the one person because of what may be the case does not show there is any greater utility being served. Again, a Utilitarian would most likely argue that one must switch the tracks towards the one person trapped in the Trolley case because it will be more likely to result in best consequences overall. The consequences Utilitarians consider, as mentioned, do not end with those directly impacted by the action, in this case those trapped on the tracks and the person deciding whether or not to switch the tracks. Recall Utilitarianism is interested in maximizing the good consequences for all people from the action. So the consequences to family members and friends of all of the individuals involved are

---

Utilitarianism can form general rules for action that tend to maximize the greater good regardless of whether a particular act may or may not lead to the greater good. Rule Utilitarians therefore may justfy switching the tracks on a different basis than Act Utilitarians. Some may even argue that the Rule Utilitarian justification would be closer to a non-consequentialist line of defense. Hence it will be more to our purpose (of contrasting the two) to assume here an Act Utilitarian justification.

also being morally considered as well as larger impacts to society in general. Everyone's interests count in determining whether switching the tracks was morally right or wrong. All other things equal, then, killing three will be worse than killing one, and we are assuming in the Trolley Problem that all other things are equal.

Part of the explanatory strength of Utilitarianism lies in its ability to simplify certain moral problems by focusing on the actual results of the action, as illustrated by the Trolley Problem. A Utilitarian can more readily justify choosing to save three lives instead of sparing only the one because the consequences of three lives being saved tends to maximize utility overall. The greater good is better served in general by trying to save more lives.[40] Non-consequentialist theories, on the other hand, have a more difficult time justifying the same choice of saving three lives over one life. Dilemmas such as the Trolley Problem bring out both the strength and weaknesses of both consequentialist theories such as Utilitarianism and non-consequentialist theory approaches to moral problems. But before turning in a moment to our examination of non-consequentialist views it is worth noting and eventually coming back to the fact that Prospect Theory will probably have implications for Utilitariainism. Given Utilitarianism's focus on maximizing positive consequences, one can already see how Expected Utility Theory is similar to Utilitarian moral decision-making. Then, one obvious reason Prospect Theory has implications for moral philosophy is because

[40] Again, we are still assuming the restrictions present in the original hypothetical where our knowledge of other facts about the case are not disclosed to us. For more on this and related limitations see notes 31, 32, and 33.

Expected Utility Theory has a counterpart in moral philosophy, namely, Utilitarianism. Recall, Expected Utility Theory argues that when making a decision one should choose the option that has the greatest expected utility. This is essentially the consequentialist argument from Utilitarianism in moral theory. Applying expected utility reasoning in general to moral reasoning, the morally correct action would be the action that has the greatest expected utility.[41] Utilitarianism in fact argues that if utility is measured in terms of happiness or satisfaction or the like, then the morally correct act is the act that results in the greatest expected utility. While there are certain aspects of Expected Utility Theory that apply primarily to economic decision models, the essential underlying explanation for why a decision maker should make one choice over another is on the similar grounds of greatest expected utility. In later sections I will return to this point about the impact of Prospect Theory on Utilitarianism. And what emerges is that, really, in the end it's not clear whether Prospect Theory is in its essentials that different from Utilitarianism.

We turn now to the contrast with non-consequentialist theories, where the challenge from Prospect Theory seems significant and deep.

[41] John C. Harsanyi, who won the Nobel Prize in Economics in 1994, has done extensive work to establish a link between Expected Utility Theory and Utilitarianism. For more on this see Mongin 1997, p. 5.

3.3 Issues for Utilitarianism and Consequentialism raised by the intuitive basis of non-consequentialist thinking

There is some good intuitive strength to the Utilitarian account of morality since in general we are inclined to say that the reason wrong actions are wrong is that they do lead to bad consequences and right or good actions lead to good consequences especially over the long run. However, critics have raised several issues with Utilitarianism as a moral account with many of the criticisms turning on the fact that things other than the consequences of an act must be taken into account before deciding on the act's character.

Non-consequentialists argue, against Utilitarianism, that what is wrong with an act of killing must consist precisely in these other things—the intrinsic character of the act, or aspects of the agent's character or intent, rather than just the consequences. Accordingly, a reliance on the consequences alone for moral permissibility not only does not get at the heart of what is truly morally problematic about impermissible actions but results in morally approving actions that many would find morally impermissible and even reprehensible. An issue raised for Utilitarian theories for example is that one instance of killing could result in more good consequences than negative consequences and would require morally justifying killing in some cases that we may consider morally reprehensible or abhorrent. A famous example that Phillipa Foot cites is an example of killing an innocent person in order to quell an angry mob

which she contrasts with the trolley problem (or runaway tram, as she calls it) previously discussed Foot says,

> Suppose that a judge or magistrate is faced with rioters demanding that a culprit be found for a certain crime and threatening otherwise to take their own bloody revenge on a particular section of the community. The real culprit being unknown, the judge sees himself as being able to prevent the bloodshed only by framing some innocent person and having him executed. Beside this example is placed another in which a pilot whose aeroplane is about to crash is deciding whether to steer from a more to a less inhabited area. To make the parallel as close as possible it may rather be supposed that he is the driver of a runaway tram which he can only steer from one narrow track on to another; five men are working on one track and one man on the other; anyone on the track he enters is bound to be killed. In the case of the riots the mob have five hostages, so that in both the exchange is supposed to be one man's life for the lives of five. The question is why we should say, without hesitation, that the driver should steer for the less occupied track, while most of us would be appalled at the idea that the innocent man could be framed. [42]

Killing the innocent person to quell the angry mob may very well result in overall greater utility since more innocent lives are in danger of being killed by the angry mob, especially if the judge frames the innocent person. A Utilitarian may therefore be forced to reason that this act of intentionally killing an innocent would be morally permissible and perhaps even morally obligatory because of the overall good consequences in terms of lives saved, happiness etc.[43] This runs contrary to many

---

[42] Foot 2002b, p. 23.

[43] Some Utilitarians may argue that even if we assume that no one ever finds out that the person executed is innocent and more lives are saved from killing the innocent person, in the long run such actions will not result in the greater good and could result in more negative consequences overall. Perhaps a Utilitarian would argue that this might lead to problems in the justice system, etc. A Rule Utilitarian, for instance, may need to create a rule that prohibits knowingly convicting the innocent since even

people's moral intuition, however. We are generally unwilling to intentionally put at risk and sacrifice the life of an innocent person even if it means more innocent people may die as a consequence. But, as Foot points out, we are willing to kill the one innocent person when they were trapped on the trolley tracks. If utilitarian consequentialist reasoning was the basis for our choice in the trolley problem, then we should similarly be willing to sacrifice the innocent person for the angry mob but we are not. This does not mean that the consequences do not matter to us in both of these scenarios but it does indicate that something other than just the immediate consequences of our actions make a difference in what we believe is morally permissible.

Nonconsequentialist perspectives on both problems may offer the necessary insight into our intuitions in both cases. A nonconsequentialist tends to argue both that we should switch the tracks in the Trolley Problem and that we should not sacrifice the life of the innocent person in the angry mob case. The non-consequentialist for instance, in contrast to a generalized Utilitarian view, would argue that the act of killing the innocent person to quell the angry mob would still be wrong even if the killing does result in a net gain in lives saved. As we have indicated

---

if the act is good overall in one particular situation, or the act is good in the short-term it will not ultimately result in the greater good overall. However, we are considering here a classical act utilitarian approach to the problem to illustrate the basic difference in perspectives of a consequentialist and a nonconsequentialist. It is also worth noting that a Rule Utilitarian response to this angry mob dilemma runs the risk of being more nonconsequentialist in its reasoning due its use of rules to help govern moral decision-making over immediate consequences. See also note 38 on Act versus Rule Utilitarianism.

in our earlier discussion of the Trolley Problem, our initial take on the reasons for why we would choose to steer the trolley towards the one seems to give a Utilitarian or consequentialist justification. But what we seek is an appreciation of the much more difficult to justify non-consequentialist, and thus non-Utilitarian, thinking in the Trolley case (in order to assess the contribution of Prospect Theory to the issue). So let us turn to that now by considering the contrasting non-consequentialist perspective on normative moral theory to that of consequentialist perspective. Then we can continue to the famous non-consequentialist principles of Doing and Allowing and Double Effect, both of which have been challenged by Prospect Theory.

3.4 The Doctrine of Doing and Allowing: A Fundamental Non-Consequentialist Ethical Principle

As already noted, non-consequentialist theories, in contrast to consequentialist theories, determine moral correctness based primarily on qualities of the action in and of itself,and not simply the consequences. The consequences of an action are still significant for a non-consequentialist, but the consequences *are not the only reason* for a moral action being considered to be permissible or impermissible. A non-consequentialist theory could argue for example that killing the innocent is wrong because it does not respect them as persons or respect their intrinsic value as human beings. Therefore the act of killing innocents is in general morally wrong regardless of whether good or bad consequences actually result from any particular

act of killing. So already we can see why a nonconsequentialist may refuse to choose to kill the innocent person for the angry mob since such a killing in general is wrong in itself. Though what is not yet clear is how a nonconsequentialist could justify switching the tracks in the Trolley Problem. This has been a very difficult problem for non-consequentialists, and Foot's and (especially) Quinn's solutions to it are generally considered to be the most compelling. In order to understand why, we must first lay out two exceedingly influential non-consequentialist principles, beginning with the Doctrine of Doing and Allowing.

The Doctrine of Doing and Allowing is generally considered to be one of the most intuitive and fundamental non-consequentialist principles. The basic idea of the Doctrine of Doing and Allowing is that there are harmful actions that it is permissible to allow to happen that one could not actively bring about or do.

In other words, the Doctrine of Doing and Allowing as usually interpreted says that it is generally worse to do something bad or harmful than it is to allow it. Non-consequentialists believe that the principle holds even when it is in violation of the principle of utility or maximization of good states of affairs. For instance in the mob case, a non-consequentialist could argue that it is not permissible to sacrifice the innocent person by framing and executing him for the supposed 'greater good'. This is in line with our intuitions in the mob case where we feel it is morally wrong to offer up an innocent person even if it means risking some greater harm. The trouble in the Trolley case, as Warren Quinn, Phillipa Foot, and many others have discussed, is

that the Doctrine of Doing and Allowing would seem to forbid our pulling the lever to move the Trolley onto the alternative track rather than allowing it to continue on its current path killing three. And yet almost everyone, including Quinn and many other non-consequentialists, finds this result intuitively unacceptable. It seems we ought to pull the lever even though that also seems to be doing harm rather than allowing harm to be done.

What Quinn argues is that the act of allowing something to occur is not always morally different from performing the act which causes it to occur. Choosing to 'allow' the trolley to continue on the tracks, killing the three, can be morally equivalent to a 'doing,' i.e. morally equivalent to a choosing to switch the trolley onto the other tracks in order to kill the one trapped there instead.

Quinn's argument will depend also on another traditional, and widely employed, non-consequentialist distinction: namely that found in the Doctrine of Double Effect. Let us get the background for his argument and these and other distinctions he introduces, by considering first Foot's original presentation of the Doctrine of Double Effect and the Doctrine of Doing and Allowing in the context of her famous attempt to address issues raised by the Trolley Problem.

3.5 Foot's Rescue Dilemmas I and II: Doing and Allowing Illustrated

For this we must turn to two other cases Foot introduced into the discussion: namely Rescue Dilemmas I and II. In Rescue Dilemma I, according to Foot,

> …we are hurrying in our jeep to save some people – let there be five of them- who are imminently threatened by the ocean tide. We have not a moment to spare, so when we hear of a single person who also needs rescuing from some other disaster we say regretfully we cannot rescue him, but must leave him to die.[44]

In considering Rescue Dilemma I a consequentialist can more readily argue that the correct action is in fact to save the five lives despite leaving the one to die. A consequentialist could defend this view on the basis that the greater amount of lives would be saved by saving the five, which in turn would maximize good consequence, all things equal. Non-consequentialist theories, on the other hand, cannot argue along these lines and instead must appeal to something about the action in and of itself to justify the same action of saving the five. In general non-consequentialists can argue one should save lives by saying that we have a duty to preserve life. In this case, if there is only a choice between saving one life and five lives then it is morally preferable to save more lives, given the duty to preserve the lives of all of the individuals at risk of drowning. Note that the reason a non-consequentialist has for saving their lives is because of the duty to preserve life out of respect for the intrinsic value of the life and not merely as a means to the end of maximizing positive consequences. The reason the non-consequentialist sees it as a duty to save the five over the one is due to the equal force of each individual's right to have their life preserved by being saved. Let me emphasize that the reader need not have the same intuition (some of my readers may be consequentialists). I only wish to make the

[44] Foot 1984, p. 179.

differences clear, and the grounds for the differences as generally accepted by non-consequentialists.

The special problem for non-consequentialists in this context arises when considering Foot's Rescue Dilemma II. Here Foot says:

> …we are again hurrying to the place where the tide is coming in in order to rescue the party of [five] people, but this time it is relevant that the road is narrow and rocky. In this version the lone individual is trapped (do not ask me how) on the path. If we are to rescue the five we would have to drive over him. But can we do so? If we stop he will be all right eventually: he is in no danger unless from us. But of course all five of the others will be drowned.[45]

Note that both in Rescue Dilemma I and Rescue Dilemma II we have five lives being weighed against one life but the circumstances have now changed. Again, if this dilemma is examined from a consequentialist perspective, there seems to be more reason to choose to save the five at the cost of losing the life of the one. Consequentialists can appeal to the greater good that is served by saving more lives as was done for Rescue Dilemma I and defend the choice of killing the one since that will maximize the number of lives saved.

[45] Foot 1984, p. 179.

Foot's Rescue Dilemma I:

"…we are hurrying in our jeep to save some people – let there be five of them-who are imminently threatened by the ocean tide. We have not a moment to spare, so when we hear of a single person who also needs rescuing from some other disaster we say regretfully we cannot rescue him, but must leave him to die."[46]

Foot's Rescue Dilemma II:

"…we are again hurrying to the place where the tide is coming in in order to rescue the party of [five] people, but this time it is relevant that the road is narrow and rocky. In this version the lone individual is trapped (do not ask me how) on the path. If we are to rescue the five we would have to drive over him. But can we do so? If we stop he will be all right eventually: he is in no danger unless from us. But of course all five of the others will be drowned."[47]

However, most non-consequentialists and even some consequentialists believe (all things equal) it is morally impermissible to save the five in Rescue Dilemma II by driving over and killing the one. Non-consequentialists often use this moral intuition as a reason to reject consequentialist reasoning since it seems that, all things being

[46] Foot 1984, p. 179.
[47] Foot 1984, p. 179.

equal, consequentialists do not have sufficient reason to prohibit the killing of the one in cases like Rescue Dilemma II (even though killing him is intuitively reprehensible to many). The problem though for non-consequentialists is that they must provide sufficient reason for saving the five and not coming to the aid of the one person in danger of drowning in Rescue Dilemma I, *while at the same time* providing good reason to not save the five and refusing to drive over the one person trapped on the road in Rescue Dilemma II. If non-consequentialists in general maintain that there is a duty to preserve life in both cases there needs to be an explanation that is morally consistent in both cases and defends the non-consequentialist moral view. The Doctrine of Doing and Allowing is able to provide an adequate response to this part of the problem which will be detailed at length later. But the basic idea is already clear: we are allowed to let some harms occur (death) that we are not allowed to actively bring about (by running roughshod over someone in the road, say).

However, there is also the further complication presented in the Trolley Problem of pulling the lever in order to switch the tracks away from the five and towards the one person trapped on the tracks. Pulling the lever does not intuitively seem to be an act of allowing something to occur but an actual act of doing. This makes it appear that the Doctrine of Doing and Allowing could not be the basis for a non-consequentialist defense of the choice to pull the lever and switch the tracks in the Trolley Problem. This shows why Foot's original Trolley problem poses perhaps the deepest problems for non-consequentialists like herself, and shows that there must

be appeal to more than just one principle behind the non-consequentialist intuitions across relevant cases.

3.6 The Doctrine of Double Effect: Another Fundamental Non-Consequentialist Principle

As we will see, non-consequentialists must rely at least in part on one other well known and fundamental principle: namley, the Doctrine of Double Effect[48]. The Doctrine of Double Effect expresses the idea that, in some cases, one can bring about harm through oblique, or indirect, intention that one could not bring about through direct intention. So, for example, if one kills the innocent person to quell the angry mob, one would directly intend for the one person to die in order to quell the angry mob. This is in contrast to the the case of the jet plane out of control that you must steer into the least crowded area; in that case the death of the innocent there is only obliquely intended as a foreseeable consequence of one's action but not directly one's intent. This illustrates the basic idea of Double Effect. But it will be good to consider various cases in order to get a working sense of the difficulties in stating the doctrine correctly. One of the main sticking points has to do with the moral

[48] The Doctrine of Double Effect and the Doctrine of Doing and Allowing are the terms used by Warren Quinn. See Quinn 1993. However, in the literature the same concepts are sometimes referred to as the Principle of Doing and Allowing, the Principle of Double Effect, as well as the Doing and Allowing distinction and I imagine perhaps others. For clarity and consistency though I will be using Quinn's use of "Doctrine" for both.

significance of the distinction between direct and oblique intention. Oblique intention is not always less wicked than direct.

Two cases that Foot uses to explain this are the wicked merchants and unemployed gravedigger cases. In the wicked merchant case Foots has us imagine some "wicked merchants selling, for cooking, oil they knew to be poisonous and thereby killing a number of innocent people."[49] The wicked merchants selling poisonous cooking oil can foresee the deaths of innocent people through the consumption of the poisoned oil but they do not strictly intend for the people to die. The merchants are primarily intending to sell the poisonous cooking oil.

This wicked merchant case can be compared and contrasted Foot says with

> …some unemployed gravediggers, desparate for custom[ers], who got hold of this same oil and sold it (or perhaps *they* secretly gave it away) in order to create orders for graves. They strictly (directly) intend the deaths they cause, while the merchants could say that it was not part of their *plan* that anyone should die.[50]

Both the merchants and the gravediggers would be considered, Foot says, from both legal and moral perspectives as murders. However, only the gravediggers strictly intend for the people to die in order so that they can create orders for graves. The difference is that the harm from the gravedigger's actions is directly intended but the harm of people dying from the selling of poisonous cooking oil by the merchants is only obliquely intended. The merchants could claim that it was not directly their intent to have people die even though they could foresee this consequence. In this

[49] Foot 2002a, p. 22.
[50] Foot 2002a, p. 22.

instance there is no significant moral difference as both are morally reprehensible but it sometimes does make a moral difference whether an action is merely foreseen or directly intended. In other words, for the Doctrine of Double Effect to be morally important the distinction between what is intended versus foreseen needs to sometimes make a morally significant difference but this does not require that direct intention always be considered to be morally worse than foreseeing.[51]

By contrast here are some cases where Double Effect *does* seem to work well: e.g. in case of certain abortions to save the mother and in so-called strategic vs. terror bombing cases. To take the latter case first, in war, you may be justified in taking out a munitions factory via a strategic bombing, for example, even though you know it will result in the death of some innocent bystanders. This is to be contrasted with a terror bombing of the innocent directly.[52] Another example can be seen in certain acts of abortion. An example Phillipa Foot gives is a hysterectomy that results in the foreseeable death of the fetus though the direct intent of the surgeon is to remove the uterus. Since the direct intent is to remove the uterus, this abortive act may be viewed as permissible since the harm to the fetus is only obliquely intended. Another surgery may not, however, be viewed as permissible, if it involves directly intending to harm to the fetus. In a craniotomy case, where the fetus' head must be crushed in order to

---

[51] There are several issues raised with the doctrine of double effect that will not be specifically addressed here. Quinn, however, does feel he has addressed the major concerns with the original conceptions of the Doctrine of Double Effect in his version of the doctrine.

[52] For more on this see Quinn 1989b, p. 177.

save a laboring mother's life, the direct intent would be to kill the fetus and the mother's death would be the forseeable consequence of failing to perform the craniotomy.[53]

What we will show is how Foot and then especially Quinn gather the basic insights of these cases together with those illustrating Doing and Allowing in order to develop a sophisticated non-consequentialst position on the Trolley and related problems.

3.7 Foot's Traditional Account of the Doctrine of Doing and Allowing and the Trolley Problem

The Doctrine of Doing and Allowing, as presented by Foot, attempts to address problems found in the Doctrine of Double Effect and finally gives what many consider to be an adequate explanation and defense of the non-consequentialist position in Rescue Dilemmas I and II as well as the Trolley Problem. Recall that the basic idea of the Doctrine of Doing and Allowing is that there are harmful actions that it is permissible to allow to happen that one could not bring about or do.

Foot was primarily focused on the Trolley Problem as a way to critically examine the Doctrine of Double Effect,[54] under which it is argued that one can, in

[53] Foot 2002a, p. 20-21.
[54] As Foot states "The words 'double effect' refer to the two effects that an action may produce: the one aimed at, and the one foreseen but in no way desired. By 'the

certain circumstances, be justified in knowingly causing through oblique or indirect intention the death of innocent people— an action one is not permitted to directly intend. She was originally interested in certain issues, just mentioned, around abortion and also so-called strategic bombing in a just war.

A very important point Foot makes along the way is worth quoting in full :

> …the distinction between what one does and what one allows to happen is not the same as that between direct and oblique intention. To see this one has only to consider that it is possible *deliberately* to allow something to happen, aiming at it either for its own sake or as part of one's plan for obtaining something else. So one person might want another person dead and deliberately allow him to die. And again one may be said to *do* things that one does not aim at, as the steering driver [in the trolley problem] would kill the man on the track. Moreover there is a large class of things said to be brought about rather than either done or allowed where either kind of intention is possible. So it is possible to *bring about* a man's death by getting him to sea in a leaky boat, and the intention of his death may be either direct or oblique.[55]

So to be a bit more precise we can state the two principles this way:

---

doctrine of double effect' I mean the thesis that it is sometimes permissible to bring about by oblique intention what one may not directly intend" Foot 2002a, p. 20.

[55] Foot 2002a, p. 25.

Doctrine of Doing and Allowing: there are instances where it may be permissible to allow by inaction a harm to befall someone that one foresees will occur, even though one morally could not actively bring that same harm about by one's action.[56]

Doctrine of Double Effect: the view that it is sometimes permissible to bring about by oblique intention what one may not directly intend.[57]

This will serve eventually for the purpose of comparisons with Prospect Theory, but it ought to be said that it does not do justice to the Byzantine details that arise around proper definitions of these principles.

In any case, sticking to standard examples, according to the principle of Doing and Allowing, it is frequently viewed as permissible to remove life support from a terminal patient even though giving the same patient a lethal injection may not be considered permissible, despite the same goal of relieving suffering in each case. This is distinguished from the Doctrine of Double Effect, which makes a moral distinction between foreseeable harm and intentional harm. For instance in a pair of examples used by Foot,

[56] I use "harm" here though other authors such as Quinn use the term "evil."
[57] Foot 2002a, p. 20.

> We are about to give a patient who needs it to save his life a massive dose of a certain drug in short supply. There arrive, however, five other patients each of whom could be saved by one-fifth of that dose. We say with regret that we cannot spare our whole supply of the drug for a single patient… . We feel bound to let one man die rather than many if that is our only choice. … . We can suppose, similarly, that several dangerously ill people can be saved only if we kill a certain individual and make a serum from his dead body. (These examples are not over-fanciful considering present controversies about prolonging the life of mortally ill patients whose eyes or kidneys are to be used for others.) Why cannot we argue from the case of the scarce drug to that of the body needed for medical purposes? Once again the doctrine of double effect comes up with an explanation. In one kind of case but not the other we aim at the death of an innocent man.[58]

In the case of the scarce drug we foresee that the single patient who needs the whole supply of the drug will die as a result of us giving the scarce drug to five other patients but we do not intend his death. The patient's death is foreseen but in no way desired. One way to test this is to see that if by luck more of the scarce drug were to be found we would be glad to give it to the patient who needed the larger supply of the drug and save his life. Contrast this with the second case where we must kill a particular person to make a serum from their body to save several people. Here we would not be permitted to kill the one since his death would not just be merely foreseen but would be our direct intent. This can be tested by seeing that the person's life cannot be saved because his death is a necessary means to our ends of the serum such that if their life were to be saved we would also fail to achieve the desired outcome of the serum that saves the five. Looking at these two cases one could be permitted in the scarce drug case to bring about by oblique intention the death of a

[58] Foot 2002a, p. 24-5.

patient to save lives that one could not bring about by direct intention in the serum case: that is, to directly intend the death of the patient in order to save other's lives. It is not a morally ideal situation to have to give the scarce drug to the five instead of the one, since we foresee the potential harm to the patient, but their death is in no way desired, but killing a person for his blood to save other patients with the serum is like framing the innocent in the mob case, or killing the infant in the craniotomy case. The killing is direct and cannot be conceived of as an unwanted merely foreseeable side effect.

3.8 Quinn on Doctrine of Doing and Allowing and Doctrine of Double Effect

With a working understanding of the Doctrine of Double Effect and the traditional Doctrine of Doing and Allowing and the nonconsequentialist perspective they attempt to justify, we can now finally proceed to the solution Quinn developed for the original Trolley problem. Warren Quinn in "Actions, intentions, and consequences: The Doctrine of Doing and Allowing"[59] reformulates the Doing and Allowing distinction and argues that his version of the distinction is morally significant in how it improves upon earlier formulations, including Foot's. One issue Quinn raises from the beginning is that there are special cases of inaction that should

[59] Quinn 1989.

be morally considered as 'doings' and, similarly, that there are special cases of action that should be seen as 'inaction'. If so, we cannot accept the distinction between doing and allowing as simply the difference between action and inaction, to do so fails to adequately capture our moral judgments.

Quinn maintains, based on another case that we will consider in a moment, that the crucial difference when considering cases of harming is between harmful negative agency, i.e. harmful refraining from an action, and harmful positive agency, i.e. a harmful active performance of an action. Harmful positive agency is "harm occurring because of what the agent does (because of the existence of one of his actions)" and harmful negative agency is a "harm occurring because of what the agent did not do but might have done (because of the noninstantiation of some kind of action he might have performed)."[60]

Quinn's Harmful Negative Agency: "harm occurring because of what the agent did not do but might have done (because of the noninstantiation of some kind of action he might have performed)."

Quinn's Harmful Positive Agency: "harm occurring because of what the agent does (because of the existence of one of his actions)."

[60] Quinn 1993a, p. 156.

Quinn notes that Phillipa Foot in her article "Morality, Action and Outcome"[61] anticipates this point, i.e. that the distinction between positive and negative agency is not the same as the difference between action (doing) and inaction (allowing). Foot, however, believes the crucial distinction lies, as she says, in the "relation an agent can have to a sequence of events that leads to harm."[62] So, in her view, negative agency is to be understood as *the allowing of a sequence of events to complete itself* and positive agency as the *maintaining or initiating of a sequence of events*.

Foot's Negative Agency: the allowing of a sequence of events to complete itself.

Foot's Positive Agency: the maintaining or initiating of a sequence of events.

Quinn argues, however, that Foot's distinction still does not adequately capture the distinction relevant to the principle of Doing and Allowing. First, he notes that Foot's account has the difficulty of determining what completes a sequence versus what initiates or continues a sequence. More important, there is also an issue, as he indicates, with certain types of action (doings) that do not seem to involve maintaining, initiating, or preventing any sequences of events. Here, Quinn appeals to a hypothetical case of firing up a neighbor's furnace.[63] The neighbor is old and might freeze if his furnace runs out of fuel. You have always helped the neighbor by

[61] Foot 2002a.
[62] Quinn 1993a, p. 159.
[63] Quinn 1993a, p. 160.

firing up his furnace though you have never promised to do so. However, five equally close friends who live far away have an emergency and need you to save them. Yet this would mean being unable to add fuel to your neighbor's furnace thereby most likely leading to your neighbor freezing to death. It is important that Quinn has us suppose that the neighbor was not in danger of freezing until this emergency. On that supposition it seems Foot's understanding of the difference between negative versus positive agency is an inadequate moral guide. Again, for Foot negative agency is to be understood as the allowing of a sequence of events to complete itself and positive agency is to maintain or initiate a sequence that leads to harm. However, in the neighbor's furnace case, not helping the neighbor is neither maintaining a sequence that leads to harm nor allowing a sequence of events to complete itself. After all, the neighbor has not begun freezing and in fact never had been freezing. But then it seems, as Quinn suggests, that one cannot justify leaving the neighbor to eventually freeze by simple appeal to the fact that one never promised to fire up the furnace and that it is only an "allowing" of the neighbor's death. Certainly in an ordinary sense of allowing, such a justification would be suspect. Clearly, then, as Quinn wishes to point out, there needs to be a better account of negative and positive agency in the Doctrine of Doing and Allowing than what Foot believed was necessary. The distinction of doing and allowing is necessary but the account of negative and positive agency must go further than saying it is the

difference between, on the one hand, allowing a sequence of events to complete itself, or maintaining or initiating a causal sequence, on the other.

It is worth emphasizing, again, that one may not share all of Quinn's intuitions here but it is not necessary for our purposes that his view be ultimately convincing. We are considering some of the most discussed moves in the debate between deontology and consequentialim on the issues raised by the Trolley Problem, the Doctrine of Doing and Allowing, and, as we will see, Prospect Theory. I want to give as much of the subtleties of Quinn and Foot's positions as I can, in order for the reader to appreciate the plausibility and argumentative force of the position. But my main concern is to eventually set up what is necessary for our critical assessment of the challenge from Prospect Theory, rather than to defend to the hilt the non-consequentialist positions developed by Foot and Quinn. What is important is that one can see a plausible line of moral reasoning from a non-consequentialist point of view.

Neighbor's Furnace Case

Background: Neighbor is old and might freeze if his furnace runs out of fuel and you have always helped the neighbor by firing up his furnace though you have not promised to do so

Decision Options
Option 1: Fire up neighbor's furnace and do not help five equally close friends who need you to save them

Option 2: Do not fire up neighbor's furnace, which leaves him in danger of freezing, in order to save five equally close friends

In rejecting Foot's account Quinn turns to his own account by first looking at cases involving the 'action of objects or forces' where the agent has some control over the objects or forces such as in the case of the trolley[64]. He shows how these and similar cases require a 'more complex treatment' than simple employment of the distinction between action and inaction. In showing this, Quinn discusses Foot's Rescue Dilemmas I and II but for now we are mostly interested in Quinn's own

[64] See Quinn 1993a, p. 160.

Rescue Dilemma III and Rescue Dilemma IV, which play a crucial role in Quinn's correction of Foot's position.

Rescue III is a slightly altered version of Foot's Rescue II and, importantly similar to the Trolley Problem. So, in Rescue III, we are to imagine that you can affect a moving train en route to save five people. You need not do anything to keep the train going on its mission to save five people, but you *can* do something to stop the train, you have control of that. On this journey to rescue the five people, you notice a person is on the track and will be killed by the train unless you act. If you stop the train, you will save the one person but will be unable to save the five people in time. Quinn argues that to continue on your mission to save the five is the wrong choice since you can stop the train and save the one person and it makes no moral difference whether you were the one who started the train, who is driving the train, or who could just access the train brake. The important moral facts in this scenario are that you *can* stop the train, and that if you don't stop it is because you intend for it to continue on over the spot where the man is stuck on the tracks. Recall again that Quinn is aiming for a non-consequentialist solution to this and related Trolley-like Problems and that a more simple consequentialist appeal to saving more lives by killing the one person and continuing to save the five is not sufficient. This would be justifiying the action via an appeal to the consequences of the action and not to something about the character of the action itself. Non-consequentialists in general cannot easily justify (if at all) directly killing one innocent person in order to save

more lives. Quinn focuses on the fact that in Rescue Dilemma III you are able to stop the train, and that you intend for the train to go forward over the man on the track; so when you allow the train to continue, you are involved in an instance of positive harmful agency.

Quinn's Rescue Dilemma III (a variant of Foot's Rescue II):

Scenario: Quinn says "We are off by special train to save five who are in imminent danger of death. Every second counts. You have just taken over from the driver, who has left the locomotive to attend to something. Since the train is on automatic control you need do nothing to keep it going. But if you do stop, and then free the man, the rescue mission will be aborted."[65]

Option One: Stop train to save one person trapped on tracks, which aborts the rescue mission resulting in failure to save five.

Option Two: Let train automatically continue on tracks killing the one person trapped on the tracks in order to complete rescue mission to save the five people.

It is important to emphasize that the fact that you can stop the train is not the only important aspect of Rescue Dilemma III for Quinn; it is important too that you intend for the train to go on right over the man on the track for the sake of the five.

[65] Quinn 1993a, p. 160.

The importance of both features of the case is illustrated by Quinn's Rescue IV, again and importantly both a variant of the original problematic Trolley and of Foot's Rescue I. In Quinn's Rescue IV he says you are (again) on a moving train where there has been an explosion injuring five passengers. As before, you can stop the train, but this will cost precious time, so you switch the train to automatic and go to help the five critically injured by the explosion. You are taking care of the five when you discover that there is one person stuck on the track who will be killed should the train continue along the tracks. In this case Quinn's judgment is that it is permissible to stay with the five instead of saving the person trapped on the tracks, and if the judgment is correct the case shows the importance of one's intentions in both Rescue Dilemma III and Rescue Dilemma IV. In Rescue III you intend the train to continue on the track, which results in the trapped person's death, whereas in Rescue IV you "intend no action of the train that leads to the man's death."[66] For instance, in Rescue IV your ends would still be fulfilled even if the train brakes were to be accidentally applied saving the one person trapped on the tracks.[67]

---

[66] Quinn 1993a, p. 162.

[67] Some may object that it is also true in Rescue Dilemmas II and III that you intend no action of the train that leads to the man's death because you'd be quite alright if by some miracle the train hit a bump and jumped over the man and continued back on the track towards the five. This miracle test may just show the need for a strict hypothetical but does not run counter to our intuitions about the original case. There has to be some constraint on what can be used within a hypothetical to test one's direct intention. Otherwise you could say in the craniotomy case that you do not directly intend the fetuses death because you'd be OK if the fetus miraculously survived the fatal procedure (by God somehow keeping the fetus alive throughout the proceedure and then restoring his skull or some such).

We see here how Quinn combines the insights of his improved Doing and Allowing distinction and the Doctrine of Double Effect (which focuses on intentions).

Quinn's Rescue IV (a variant of Foot's Rescue I):

Scenario: Here Quinn says "…you are on a train on which there has just been an explosion. You can stop the train, but that is a complicated business that would take time. So you set it on automatic forward and rush back to the five badly wounded passengers. While attending to them, you learn that a man is trapped far ahead on the track. You must decide whether to return to the cabin and save him or stay with the passengers and save them."[68]

Option One: Return to the cabin and stop the train to save the man stuck on the tracks and fail to save the five badly injured passengers on the train.

Option Two: Stay with the five wounded passengers and let the train automatically continue on track failing to save the one man trapped on the track in order to save five injured people on the train.

So for Quinn in Rescue III, there is both *control* and *intention* in the agent's choice to not stop the train, and so the agent's decision is in this way a deliberate action, and therefore, morally the same as directly intending to run over and kill the one person on the tracks. Thus, for Quinn, this is a case of *positive agency*. This is an

[68] Quinn 1993a, p. 161.

instance where what we would generally call an 'allowing' or inactivity, is morally equivalent to a 'doing,' namely of intentionally running over the person on the tracks. Quinn's Rescue III and IV illustrates that what is *intended* (namely, the death of the man stuck on the track in III)—versus, what is merely *foreseen* (the death of the man stuck on the track in IV)—must also be an aspect of Doctrine of Doing and Allowing. Hence, the Doctrine of Double Effect, where foreseeability is an important aspect, plays a role in the proper application of the Doctrine of Doing and Allowing. In particular, the Doctrine of Doing and Allowing and the Doctrine of Double Effect are both necessary for a complete explanation of non-consequentialist intuitions in the Trolley Problem and its variants.

The Doctrine of Double Effect provides a moral explanation of the importance of direct versus indirect intent. The Doctrine of Double Effect therefore focuses on the purpose for the action being undertaken. So when examining a particular harm the Doctrine of Double Effect examines what the agent intended to do, his *purpose*, and the Doctrine of Doing and Allowing examines what the agent actually did or did not do, his *action*.

The Doctrine of Double Effect can recognize that the same action could be performed but with different intents. Whereas the Doctrine of Doing and Allowing can recognize that there can be a difference between actions we perform by 'doing' and actions we perform by refraining from doing an act.

Quinn does not believe the two distinctions can be collapsed into one grand distinction since the Doctrine of Double Effect depends "on whether or not a victim is *himself* an intentional object, someone whose manipulation or elimination will be useful."[69] This is in contrast to the Doctrine of Doing and Allowing, where there is no assumption that the victim herself is an intentional 'object' who is useful to 'manipulate or eliminate'. In Rescue III, for instance, the victim who is trapped on the tracks is an intentional object whose death is intended in order to save the five people. The trapped person's death is seen as an essential element in the saving of the five people. In Rescue IV, by contrast, you have control over whether to stop the train from hitting the trapped person, but it is not part of your direct intent that the train continue forward. As Quinn says, "In Rescue III, but not in Rescue IV, the train kills *because* of your intention that it continue forward."[70] The Doctrine of Double Effect therefore not only discriminates generally against harmful direct agency, it also addresses the fact that people "may figure as intentional objects not only of a choice to act but also of a choice not to act."[71] Quinn, in other words, is saying that people can be a means to your ends, and not just a side effect of your ends, whether you are refraining from an action or engaging in an action. In this way the Doctrine of Double Effect distinguishes between what is "actively brought about and what is merely

[69] Quinn 1993a, p. 162.
[70] Quinn 1993a, p. 162.
[71] Quinn 1993b, p. 187.

allowed to happen."[72] This overlaps with the distinction of allowing harm and harming in the Doctrine of Doing and Allowing, except that the Doctrine of Double Effect sometimes "discriminates against direct agency in which harm is done. And sometimes … it discriminates against direct agency in which harm is allowed."[73]

Given the above considerations, Quinn focuses on the idea of an agent's contribution to an event as an essential factor in the relevant distinction in Doing and Allowing:

> An agent's most direct contribution to a harmful upshot of his agency is the contribution that most directly explains the harm. And one contribution explains harm more directly than another if the explanatory value of the second is exhausted in the way it explains the first.
>
> In the absence of special circumstances involving the actions of objects, an agent's contributions to various effects in the world are those of his voluntary actions and inactions that help produce the effects. So in ordinary cases, his most direct contribution to any effect is the action or inaction that most directly explains the effect.[74]

So in Rescue Dilemma IV the most direct contribution to the death of the one is not the decision to save the five, which is a less direct contribution, but rather the decision to not save the one, Harmful Negative Agency.

Later we will see Horowitz use some of these scenarios to apply Prospect Theory as an alternative analysis to the principle of Doing and Allowing. But for the moment, let us complete our review of the dilemmas and turn again to Quinn's

---

[72] Quinn 1993b, p. 187.
[73] Quinn 1993b, p. 187.
[74] Quinn 1993a, p. 163.

analysis of Foot's Rescue Dilemma I, where we can only either save one person from drowning or five from drowning—we cannot save everyone.

Quinn argues that we are justified in acting to save five people in Rescue Dilemma I since we would only be 'allowing' one person to die whom we could not save without allowing 5 to die instead. This is a simple case of doing and allowing.

3.9 Quinn's Final resolution of original trolley problem

Recall that one of the main difficulties of the original Trolley is that it does not seem that we can appeal to the Doctrine of Doing and Allowing as the basis of our decision to steer the trolley towards the one innocent person and away from the five. In Quinn's terms, the action cannot be seen as an instance of Harmful Negative Agency that is morally preferable to Harmful Positive Agency as generally indicated in a traditional Doctrine of Doing and Allowing account. In the original Trolley Problem the agent has control of the direction of the trolley but in order for there to be Harmful Negative Agency, recall, the harm must occur because of what the agent did *not* do but might have done. However, switching the tracks toward the one man trapped would cause a harm due to the action the agent *does* do, in this case, switching the tracks. Therefore, switching the tracks is clearly a case of Harmful Positive Agency. What's worse is that it does not seem that we can plead Double Effect either, since in steering the Trolley directly towards the one innocent it is

implausible to say that we do *not* intend his death directly, and only see it as a merely foreseeable side effect of what we do.

Now, we have seen that the main problem posed by the original Trolley is that the other option, to not switch the tracks, appears at first to be an instance of Harmful Negative Agency. This is due to the fact that not switching the tracks seems like an action one is refraining from, and, in general, the agent's action is passive versus the active act of switching the tracks. If switching the tracks is an action an agent could have done but did not do, then the harm to the five being struck and killed by the trolley seems to be Harmful Negative Agency. This is why the original trolley problem is used as a counterexample to the Doctrine of Doing and Allowing, as traditionally understood (before Quinn's analysis) since it seems clearly wrong to all to allow the trolley to continue on its course to kill the five and yet allowing the trolley to continue seems to be what is indicated on a traditional Doctrine of Doing and Allowing account. If the traditional Doctrine of Doing and Allowing account consistently disfavors Harmful Positive Agency and favors Harmful Negative, then it seems non-consequentialists cannot say it is permissible to switch the tracks to save the five.

Quinn argues, however, that choosing to not switch the tracks, just like choosing to switch the tracks, is actually Harmful *Positive* Agency. For Quinn, deciding to not switch the tracks towards the one is the same as directly intending that

the trolley move forward towards the five. Part of the reason for this is because of the motive behind the act of not switching the tracks. Quinn believes that the aim must be to save the one man trapped on the tracks instead of the five on the tracks. To achieve this goal of saving the one man, the trolley must continue on its course past the switch on the tracks which then kills the five. If you choose to not switch the tracks then you are directly intending to save the one, and as part of that purpose also intending for the trolley to continue forward over the five, and therefore, just as in Rescue III (where you have an analogous intent and control), you are involved in Harmful Positive Agency despite the passive appearance of the act itself.[75]

So in the end Quinn's solution to the original trolley problem is to treat both options, switching and not switching the tracks, as Harmful Positive Agency. Given that both options are Harmful Positive Agency, the trolley problem becomes a decision about which option does the least amount of harm. Switching the tracks towards the one man trapped results in less harm than choosing to have the trolley continue on its course to kill five.[76]

Notice this includes a consequentialist basis since in the end it is the number of lives that is the deciding factor for the correct moral action. Nonconsequentialists, however, can consider consequentialist reasoning as a *part* of their moral justifications. The difference for a nonconsequentialist though is that the

[75] While not switching the tracks is an inaction since you are refraining from an act, what is important here for Quinn is that not switching the tracks still involves agency (intent and control).
[76] Quinn 1993a, p.166-7.

consequences are not the *only* basis for justifying moral action. Nonconsequentialism is not the crazy view that consequences don't matter at all, they do matter, but they are not the only or primary thing. Also note that Quinn's solution to the trolley problem leaves the Doctrine of Doing and Allowing Distinction (as he amends it) intact since both acts are considered Harmful Positive Agency.

In other words, what most directly explains your action of switching the tracks is that you are redirecting the trolley away from the five in danger. This does move the trolley towards the one person endangering them so this *must also be part of your intent* and his death is also a *necessary* part of redirecting the trolley. This is similar to the Foot's Rescue Dilemma II where you must run over the one in order to save the five because there the person stuck in the path on the road is an inseperable part of your ends to save the five.[77] But of course in Dilemma II, unlike the Original Trolley, you have the option of a lesser evil in terms of the Harmful Negative Agency involved in not saving the five. You must intend the one person's death in order to save the five because you must formulate the direct intention to run over him, i.e. to

[77] Again one could object here that he is not a means to our ends here either, by pointing out that we would be happy if the train magically skipped over him on your way to the five. The serum case discussed earlier provides a closer means to one's ends case since there we need to kill the person in order to get the serum. Your intent would remain the same even if a miracle occurred saving the person through a magical skipping over in the trolley case or a different already dead body in the serum case. Contrast this with switching the tracks and killing the one where your intent does not require the death of the one in order to attain the ends of saving the five. But see note 66 above for my response to this sort of magical or miraculous hypothetical test of intent.

kill him. But you are not permitted to do this because you have the option of simply refraining from saving the five.

Quinn's Doctrine of Doing and Allowing provides an explanation that is consistent across all of these dilemmas, Foot's Rescue Dilemmas I and II as well as the Trolley Problem and others, giving non-consequentialist theories solid footing for a defense that for many is more convincing than any consequentialist explanation of our moral judgments.[78] You need not be fully convinced by Quinn's account but you should be able to see how Quinn's view offers a plausible nonconsequentialist line of reasoning that can address these dilemmas.

As we've noted above, the Doctrine of Doing and Allowing is an extremely influential doctrine in ethics since it provides a convincing explanation of our moral judgments not just in rescue dilemmas such as these, but in many other contexts in every day life. For example, in general we think it is worse to harm someone intentionally than to fail to help someone, even if failing to help resulted in some harm. Our good samaritan laws that protect those who fail to come to someone's aid who need help (and protect those who do help) seem to reflect this notion. Quinn is also generally considered to have one of the most subtle and sophisticated

[78] Recall, that nonconsequentialists need a morally consistent explanation of their intuitions for these dilemmas. The trolley problem made this more difficult since a nonconsequentialist's willingness to switch tracks, killing one and saving five, looks consequentialist. So nonconsequentialists need to have an explanation that can defend switching tracks in the trolley problem, which allows a person to die in the one case, and still defend decisions in other dilemmas, such as the decision to not run over and kill the innocent person in Foot's Rescue Dilemma II.

developments of the Doctrine of Doing and Allowing in contemporary moral theory.

4. Implications of Prospect Theory for Nonconsequentialists

Philosopher Tamara Horowitz argues that Kahneman and Amos Tversky's Prospect Theory should be the preferred explanation of moral decision-making over Warren Quinn's Doctrine of Doing and Allowing in ethical theory. Horowitz believes Prospect Theory offers a psychological and better explanation of our moral judgments.[79]

As mentioned earlier, the impact of Prospect Theory in philosophy so far is primarily in ethical theory as it has been applied specifically to *moral* decision-making. While we have discussed some potential implications for utilitarian and for consequentialist reasoning in general, for now what I want to concentrate on are the implications of Prospect Theory for non-utilitarian, and, more generally, *non*-consequentialist theories.

4.1 Application of Prospect Theory vs. the Doctrine of Doing and Allowing

We have already discussed several dilemmas and scenarios, which is understandably trying. But we must introduce yet another since it is not only a well-cited and key part of Khaneman and Tversky's research but it also forms the basis for Horowitz's argument. Horowitz in particular aims to show that the Doctrine of Doing

[79] Horowitz 1998.

and Allowing from moral theory is not necessary to explain intuitions such as Quinn's in Rescue Dilemmas III and IV or his intuitions in Foot's Rescue Dilemmas I and II. Instead, she suggests Prospect Theory offers an alternative explanation by only appealing to our decision-making psychology. Horowitz needs for her argument to show a connection between existing psychological Prospect Theory research and moral theory. The Asian Disease case is this link.[80]

4.2 Asian Disease Problem

With a basic understanding of Prospect Theory as described in our account of it in section 2.0 above, we now turn to the Asian Disease case. This case involves an experiment with two groups, where the first group is given the following hypothetical:

> Imagine that the U.S. is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Assume that the exact scientific estimate of the consequences of the programs are as follows:
> If Program A is adopted, 200 people will be saved.
> If Program B is adopted, there is 1/3 probability that 600 people will be saved, and 2/3 probability that no people will be saved.
> Which of the two programs would you favor?[81]

[80] Horowitz 1998, p. 370.
[81] Kahneman and Tversky 1981, p. 453.

For the second group of subjects, instead of programs A and B, they are given alternative programs C and D in the same scenario with the following description:

> If Program C is adopted, 400 people will die.
> If Program D is adopted, there is 1/3 probability that nobody will die, and 2/3 probability that 600 people will die.[82]

Similar to the two pairs of options in the Richer Decision problem (see chapter 2, section 2.2., pages 12 and 13), both program choices A and B and program choices C and D are identical to each other in terms of having the same expected utility. Group I's Program A will result in the same amount of lives saved, 200, as Group II's Program C. Similarly, Group I's Program B will have the same expected survival rate, 1/3 chance of 600 people surviving or a 2/3 chance of 0 people surviving, as Group II's Program D. The only difference between the options presented to both Group I and Group II is in how the expected survival rate was stated: either as "people will be saved" or "people will die."

The first experimental group, when evaluating the two programs A and B, tended to choose Program A guaranteeing that 200 people will be saved. Prospect Theory can explain this choice by first pointing out that since the program uses the phrase "people will be saved," this leads participants to view the 600 deaths as the status quo neutral reference outcome and saving lives as a gain, of 200 lives since it is a positive deviation from 600 people being dead. Program B's 1/3 probability of 600

[82] Kahneman and Tversky 1981, p. 453.

lives saved would also be a positive deviation from this neutral reference outcome but, due to the fact that it is a risky prospect, it is not valued with the same decision weight as saving 1/3 of the 600 lives for sure. Program B, recall, also contains the 2/3 probability that no one will be saved: which is neither a positive nor negative deviation and simply remains at the no gain neutral reference outcome of 600 lives lost. Group I in the end was risk averse to the possible gains in lives offered by Program B and so, instead of risking that no one would be saved by opting for Program B (in hopes of saving 600 people instead of just 200), they tended to choose Program A saving a certain 200 lives. Recall again that according to Prospect Theory, people tend to avoid negative deviations from the neutral reference outcome before trying to secure positive ones. In this case the risk is in remaining at the neutral reference outcome, a no gain, which is valued lower in comparison to the certain gains above the neutral reference outcome. In other words, Program B in the evaluation phase is lower in desirability to Program A because Program B risks no gains instead of securing with certainty the positive gain of 200 lives in Program A.

With this in mind, notice how the second experimental group evaluated their two programs C and D when their programs were phrased instead as "people will die". Again, Expected Utility Theory would anticipate that both Group I and Group II would tend to choose the same option that has the greater expected utility, since the

choice pairs for each group are the same in terms of expected lives saved.[83] Group I should choose Program A and Group II should choose the equivalent Program C for Expected Utility theory (as a theory of rational human conduct) to hold true. However, Group II tended to choose Program D instead, contrary to what Expected Utility theory indicates.

As Horowitz explains, "in the first experiment, this shift in neutral reference point [from saving lives to losing lives] presumably forces the subjects to evaluate disease fighting programs as leading to positive deviations in one case [Group I] but negative deviations in the other [Group II]… ."[84]

So Prospect Theory explains the unexpected result of Group II choosing Program D by looking again to how the experimental group frames its decision in the editing phase and in turn how this results in a different valuing of choices in the evaluation phase. Here, Prospect Theory argues that the Group II participants set their reference point at 600 people alive and not at 600 people dead as Group I had done. So even though they risk more people dying by choosing Program D, where there is a 2/3 probability of 600 dying, this is only a difference of 200 lives from the neutral decision point of 600 people alive. Furthermore, there is at least a good 1/3

[83] Since all of the program options have the same expected utility, one could argue that there should be no preference for either Program A over Program B or for either Program C over Program D. However, experimental evidence shows a clear preference of Program A over Program B and Program D over Program C. See Kahneman 2011.

[84] Horowitz 1998, p.375. Note that Horowitz here refers to "neutral reference outcome" as "neutral reference point." She also sometimes uses "neutral outcome" instead of "neutral reference outcome" as well. See also earlier note 17.

chance of no one dying so they choose the risky Program D in hopes of avoiding losses from the reference point. So, the second experimental group, instead of being risk averse like Group I, was risk seeking and choose Program D since, according to the theory, we are risking seeking when avoiding losses.

Asian Disease Problem

Description:
Imagine that the U.S. is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Assume that the exact scientific estimate of the consequences of the programs are as follows:

Group 1
Program A; 200 people will be saved.
Program B; 1/3 probability that 600 people will be saved, and 2/3 probability that no people will be saved.

Group 2
Program C; 400 people will die.
Program D; 1/3 probability that nobody will die, and 2/3 probability that 600 will die.

Prospect Theory thus can explain the difference among the subjects' choices in both the earlier Richer Decision and Asian Disease decision scenarios because "differing instructions given to subjects in the two parts of the experiment are assumed to result in differing choices of neutral reference point".[85]

[85] Horowitz 1998, p. 373.

Hence, in each of these two cases, Horowitz's view is that expected utility theory does not provide an adequate explanation for the choices made.

4.3 Quinn's Rescue Dilemmas and Horowitz

Horowitz believes the Asian Disease case in Prospect Theory can be viewed as an analogous case to Rescue Dilemmas I and II. Horowitz attempts to show that since Prospect Theory can explain the differences between groups 1 and 2 in the Asian Disease case that perhaps Prospect Theory could also explain why our intuitions in Rescue Dilemma I differ from that of Rescue Dilemma II. Quinn's argument, however, is that anyone in the cases of Rescue Dilemma I and Rescue Dilemma II who share Quinn's moral judgments also implicitly accept the Doctrine of Doing and Allowing. Horowitz, on the other hand, does not believe these moral judgments involve either explicit or implicit acceptance of the Doctrine of Doing and Allowing, but instead claims these judgments are the result of the kind of decision-making explained by Prospect Theory.[86]

Let us note, first, that Horowitz changes the Rescue cases a little in order to make them more similar to each other. Rescue Dilemma II now involves saving five potential drowning victims and killing one person because this is closer to the

[86] Horowitz 1998, p. 369.

scenario in Rescue Dilemma I than the version where we drive over and kill someone in order to save five lives endangered by some unknown means or manner. The options as Horowitz lays them out are as follows:

Horowitz's breakdown of Foot's Rescue Dilemma I:

Option 1. - (a) Save a group of five people in danger of drowning, and (b) fail to save one person in danger of drowning.

Option 2. – (c)Save one person in danger of drowning, and (d) fail to save the group of five people in danger of drowning.

Horowitz's breakdown of Foot's Rescue Dilemma II: *[rewritten as a variant of Rescue Dilemma I]*

Option 3. – (e) Save a group of five people in danger of drowning, and (f) kill one person who would otherwise live.

Option 4. – (g) Fail to save the group of five people in danger of drowning, and (h) refrain from killing one person.[87]

As mentioned, Prospect Theory, Horowitz argues, can explain both of Quinn's intuitions for Rescue Dilemmas I and II. To make her argument then she needs to show how the same ranking of preferences could also be explained by Prospect

[87] Horowitz 1998, p. 376.

Theory as well in each rescue dilemma. Keep in mind she is not trying to explain the Doctrine of Doing and Allowing Distinction via Prospect Theory but is attempting to provide an alternate explanation for Quinn's intuitions that he uses in constructing the Doctrine of Doing and Allowing.

For Rescue Dilemma I, Horowitz explains that the phrasing of Option 1 and Option 2 as "fail to save" make it seem that at least one person dying is the status quo in the editing phase. If Option 1 is chosen, five people are saved and one person dies and if Option 2 is selected one person is saved and five people die. Horowitz believes therefore that for Rescue Dilemma I it is reasonable to see one person dying as the reference point in the editing phase for Prospect Theory. So Option 1 is framed in the editing phase as resulting in a positive deviation gain of lives saved from the one person dead reference point since five people are saved in part (a). Option 1 does involve one death in part (b) as well but this is the death of one person already accounted for in the one person dead reference point so it is seen as an outcome that does not affect the decision. If each life is treated equally[88], Horowitz says, and we sum up the total lives lost or gained in parts (a) and (b)[89], then Option 1 results in the gain of five lives total from the reference point of one person dead.

---

[88] Horowitz assumes that it is psychologically realistic that each life can be valued the same morally though she understands this a "substantive hypothesis." For more on this see Horowitz 1998, p. 376.

[89] Horowitz says she is making some assumptions on how Prospect Theory would attempt to calculate options that involve conjunctives. She believes it is reasonable that Prospect Theory would calculate the value of each part of the conjunct separately and then take up the combined value of each as a sort of summation of all the values

On the other hand, Option 2 is seen as a negative deviation or loss from the reference point in the editing phase. Five lives are lost in Option 2 part (d) but again it is the additional movement away from the reference point that is significant in determining whether there is a loss or a gain for Prospect Theory. So in this sense the five lives lost in part (d) are counted as only four additional lives lost relative to the reference point of one person dead. In other words, there are still five lives lost in total but only four more lives lost (an outcome that one can affect by the decision) if one life of the five is already seen as being lost regardless of what one does. When the two options of four additional lives lost in Option 2 part (d) and the one life saved in Option 2 part (c) are combined to determine the cumulative outcome of Option 2, there is a total of three additional lives lost negative deviation from the reference point in Option 2. This leads to a strong preference in the evaluation phase Horowitz believes, upon comparing the two edited options, to clearly rank Option 1,which saves five lives relative to the reference point, higher than Option 2, which loses a total of three additional lives relative to the reference point. This ranking by Prospect Theory matches Quinn's preference for Option 1 over Option 2 in Rescue Dilemma I. But of course, the match is based on an entirely different line of reasoning from Quinn's, since Prospect Theory makes no appeal to the Doctrine of Doing and Allowing here.

---

involved. Though she does not believe this is generally valid in the evaluation of conjunctives. See Horowitz 1998, p. 377-80; especially pages 379-80.

In contrast, in Rescue Dilemma II, Horowitz believes that the reference point is one person alive since the phrasing here emphasizes "kill one person" instead of Rescue Dilemma I's "fail to save". Kill wording in Rescue Dilemma II's options presents the (at least) one person as being alive as the status quo and not as one person dead as represented in Rescue Dilemma I. This results in a different reference point being set in Rescue Dilemma II from the reference point that was set in Rescue Dilemma I. Regardless of whether one chooses Option 3 or Option 4 in Rescue Dilemma II, it appears that one life will be saved no matter what option a decision maker selects. This neutral outcome of one life saved, therefore, is the reference point for Rescue Dilemma II according to Horowitz. So in the editing phase for Rescue Dilemma II, killing one person, in Option 3 part (f), and failing to save five, in Option 4 part (g), are both seen as being negative deviations from the status quo reference point of one person alive. Option 3 part (e), however, saves five people, which is a positive deviation of four additional lives saved over the one person alive reference point. When Option 3 parts (e) and (f) are considered together, there is a total of three lives saved above the one person alive reference point. For Option 4 in part (h) one person is alive, but this is not seen as a positive deviation from the reference point since one person alive is the reference point. Yet when Option 4 parts (g) and (h) are taken up together there is a total of four lives lost relative to the reference point which is a negative deviation and is hence viewed as a loss. Rescue Dilemma II's options in the editing phase could then be seen as being framed as a

choice between losing four lives in Option 4,which is a negative deviation, and the positive deviation or gain (from the reference point) of saving three additional lives in Option 3. In the evaluation phase Option 3 then is ranked above Option 4 for Prospect Theory because of being averse to the losses on Option 4.

Horowitz argues that when Quinn compares Rescue Dilemma I to Rescue Dilemma II there are more compelling reasons for him to select Option 1 in Rescue Dilemma I than the reasons to select Option 3 in Rescue Dilemma II. But not for the reasons Quinn thinks. Rather, as apparently, a great deal of experimental evidence seem to show, the intuition Quinn and others have is due to the fact that when the gap in value between Options 1 and 2 in Rescue Dilemma I is compared with the gap in value between Options 3 and 4 in Rescue Dilemma II, the absolute value is greater, and hence more compelling, in Rescue Dilemma I then it is in Rescue Dilemma II. The perceived value gap, as Horowitz states, between "killing and sparing is larger than the gap in *v* [value] between letting die and saving"[90]. In other words, Horowitz says,

> When you analyze these two problems [in Rescue Dilemmas I and II] it may seem to you worse to kill than to let die. That is, it may seem to you that in the first problem [Rescue Dilemma I] the reason you have to spare the person's life is *more compelling* than the reason you have to save the person in the second problem [Rescue Dilemma II]. This way of putting the matter presupposes that an agent can compare the force of the reasons for doing one thing or another *across several distinct decision problems*. Prospect Theory does not contain this assumption, and it is problematic. It implies that an agent has

[90] Horowitz 1998, p. 380.

> evaluated the relevant outcomes in several different decision problems *on the same scale*. I doubt that this is true in general.[91]

So for Prospect Theory, the decision is not about killing or letting die at all in either rescue dilemma but concerns whether someone would similarly view their reasons for preferring Option 1 over Option 2 in Rescue Dilemma I as more compelling than their reasons for preferring Option 3 over Option 4 in Rescue Dilemma II.[92]

Because Quinn's intuitions can be explained by Prospect Theory Horowitz argues that they are not moral intuitions at all. According to Horowitz, it is "crucial for Quinn that the intuitions he elicits be moral intuitions, since he wants to argue that our moral intuitions both support the Doctrine of Doing and Allowing and conflict with consequentialism"[93]. Horowitz firmly believes, however, that Quinn is mistaken when he argues that there is a difference in terms of permissibility between Rescue Dilemma I and II. In fact Quinn is "mistaken in thinking that they are making a moral judgment at all."[94] Horowitz makes a case that Prospect Theory suggests that there is a reasoning process occurring, such as the editing and evaluation phases described by Prospect Theory, but that it is something that people are generally

---

[91] Horowitz 1998, p. 377.
[92] Yet, as Horowitz argues in the above quote, Prospect Theory assumes each agent is evaluating outcomes 'on the same scale,' in other words the values an individual gives the outcomes have the same weight from one decision problem to the next. It is this issue of equivalent 'scale' that could show how moral judgment is the underlying basis of the decision. Even if one accepts the Prospect Theory explanation this could provide room for a response to Horowitz from ethicists. However, for now, how Horowitz views Prospect Theory is still a key example to show how ethical theory is being put to the test, as it were, by results from other disciplines such as psychology.
[93] Horowitz 1998, p. 381.
[94] Horowitz 1998, p. 381.

unaware of. However, Horowitz argues that even though Prospect Theory may reveal an instance of reasoning that we may not be fully aware of one can 'consciously and voluntarily' go through the same form of reasoning as well.[95] Moral intuitions and principles can be unconscious as well but since Prospect Theory can be empirically tested and offers an alternative explanation to the Doctrine of Doing and Allowing Horowitz believes Prospect Theory should be preferred.

Since the intuitions Quinn uses to support the Doctrine of Doing Allowing in Rescue Dilemmas I and II can be explained by Prospect Theory, Horowitz argues philosophers' thought experiments do not provide adequate evidence in general for philosophical conclusions. Philosophers, Horowitz believes, are inventing principles like the Doctrine of Doing and Allowing in thought experiments after the fact of their judgment, which itself is based on something else entirely. Again there is a great deal of experimental evidence offered for these psychological claims, see Kahneman's recent discussion and references in *Thinking, Fast and Slow*, especially chapters 31-34.[96]

But as I hope to show, in a critical discussion of this one often cited case (the Asian Disease case) the evidence can be questioned on philosophical grounds.

---

[95] Horowitz 1998, p. 384.
[96] Kahneman 2011.

4.4 Asian Disease Decision Problem Critique: Equalization

To assess how well Horowitz applies Prospect Theory to the rescue dilemmas and the Doctrine of Doing and Allowing using insights from the Asian Disease case, the Asian Disease case itself needs to be examined in more detail. The reason for this is that upon closer examination there are significant moral distinctions that are not made in the Asian Disease case that could affect its applicability to moral theory. I'll examine the problems with the Asian Disease case setup and then it will help for our purposes to create a more morally clear Asian Disease case based on the original Asian Disease scenario. This will enable us to call into question Horowitz's view that the Doctrine of Doing and Allowing distinction could also be explained by Prospect Theory.

In the Asian Disease case posed by Kahneman and Tversky the moral values of the participants seem to be more in the foreground than in the earlier Richer Decision case. Participants in both groups of the Asian Disease case study are affected and influenced by whether their disease fighting program is seen as lifesaving or life risking. At first glance it seems that the participants in both groups should have selected the same option in their respective programs since, as we have seen in other Prospect Theory cases, the expected utility of each outcome is the same. However, even though the expected utility of each outcome has been equalized, the

phrasing of the Asian Disease problem's description and programs have not been similarly equalized as was previously thought.

Equalization is the idea that when comparing cases all factors other than the characteristic being tested need to be equal for the conclusions to be valid. For instance, in Foot's Rescue Dilemmas I and II, both cases involve a choice between either an action that would save five people at risk of drowning from the incoming tide, or an action that would save a single person who was also in life-threatening danger. In both Rescue Dilemmas you could not save all six people, you had to choose between only the two courses of action given in the problem. What was being tested in Rescue Dilemmas I and II was the means to the consequences in each case so what varied between the two cases was the means to saving the five people. In Rescue Dilemma I you save the five people by choosing to save the five instead of saving the one person in danger. This is in contrast to Rescue Dilemma II where you save the five people by choosing both to save the five and choosing to run over and kill the one person trapped on the path. If we decide it is permissible to save the five people in danger of drowning in Rescue Dilemma I but not permissible to save the five people in danger of drowning in Rescue Dilemma II, then we can more safely conclude, due to the equalization of the cases, that the reason for this difference in moral permissibility is due to the difference of means in the two cases.

Equalization also means that there cannot be vaguenesses that bear multiple interpretations of the same problem. Vagueness in a problem can cause multiple

issues but at minimum vagueness can fail to equalize the two cases for a proper comparison upon which to draw conclusions and at worst undermine the entire purpose for the comparison.

This does not, however, mean that there cannot be any vagueness within a problem set in order for two cases to be equalized. Some vagueness may even be necessary in order to compare two cases or to examine a particular concept. For instance in Rescue Dilemmas I and II, we do not know much about the people who are in danger in the dilemmas except for the circumstances that they find themselves in. Specific facts about the individuals have been intentionally left out of the dilemma so as to focus on the means and not on individual circumstances. So one could say that there is a vagueness surrounding the individual people that could be morally relevant. An example might be that if the one person knew the cure for cancer and the five others were terminally-ill cancer patients, there may be a reason to prefer the one person over the five in both Rescue Dilemmas I and II. Notice, however, that this would be a special case and these additional facts about the individuals in danger actually distract from the purpose of this particular pair of dilemmas. Furthermore, information about the people would actually create a lack of equalization between the dilemmas since the people themselves may not necessarily be valued equally. Vagueness about the people here actually helps to equalize the two problems so that we know that the reason one may be choosing one action over

another would be due to the means and not due to some different valuing of individual persons.

Kahneman and Tversky implicitly recognize the importance of equalization since many of the aspects of the Asian Disease problem have been equalized. Both Group 1 and Group 2, for instance, were presented with the same initial description of the problem. In addition, as has been mentioned earlier, Kahneman and Tversky set up the same expected utilities for each of the programs being considered by each group just as they did in the Richer Decision Problem. If expected utilities were not equalized than Kahneman and Tversky could not support their findings. Since if that were the case then a critic could simply question the results of their research by arguing that it was a difference in expected utilities that was the basis for Group 1 and Group 2's different decisions and not due to arbitrary framing effects as Prospect Theory describes.

The two sets of program choices for Group 1 and Group 2, therefore, must be, as Frances Kamm says, "essentially identical and differ only in trivial variations in their characterization that are not morally relevant."[97] If anything other than the framing of the program choices differs in what is presented to Group1 and Group 2, then Kahneman and Tversky cannot be certain that the reason for the different program choices is due to framing effects as understood by Prospect Theory. In the results of Kahneman and Tversky's research on the Asian Disease problem, we

[97] Kamm 2008, p.424.

already know that Group 1 selects Program A over Program B, and, Group 2 selects Program D over Program C. However, there are equalization issues in the Asian Disease problem which may have created potential differences in participant interpretations of programs by each group, and reveals morally relevant differences in the moral characterizations of the programs, raising issues for Kahneman and Tversky's findings.

The main equalization issue in the Asian Disease Problem appears to be one of vagueness. Vagueness that is not just troubling for equalization but also for moral clarity. In the Asian Disease Problem, there are several vaguenesses. There is vagueness in the phrasing of the description of the case, vagueness in the word choice of the program options, and finally a combined vagueness that comes as a result of not having clarity in both the description and the programs when they are considered together as a problem set by participants. These vaguenesses have resulted in the researchers inadvertently failing to equalize the two groups programs and have split the Asian Disease scenario into at least two morally distinct problems. Let me show you explicitly how vagueness and nonequalization might occur in the Asian Disease problem.

First, let's examine the problem setup of the Asian Disease case that is shared by both Group 1 and Group 2. In the Asian Disease problem description it states there are two alternative programs to "combat the disease."

Asian Disease Problem Description:

Imagine that the U.S. is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Assume that the exact scientific estimate of the consequences of the programs are as follows:

Notice that the phrase "combat the disease" can bear multiple interpretations. "Combat the disease" could mean combating the disease with a treatment for those infected with the disease. For example, with HIV infection, treatment often includes antiretroviral medications given to those infected with HIV to help prevent deaths that can come as a result of opportunistic infections from an HIV compromised immune system. However, "combat the disease" could also mean combating the disease with a cure for those infected. Penicillin, for instance, is a cure for many bacterial diseases that used to be considered life threatening such as syphilis. There are still other possible interpretations of "combat the disease". "Combat the disease" could refer to fighting the disease by preventing disease infections. An instance of this are the ongoing efforts to educate people about HIV so as to prevent the spread of HIV infection by reducing transmission of the disease. An additional possibilty for prevention that combats disease is a vaccine, such as the new HIV vaccine that is

being developed which would "combat the disease" by preventing an HIV exposed person from developing an HIV infection. [98]

As we can see, "combat the disease" is a vague phrase that could be interpreted in a variety of ways. Participants in Group 1 and Group 2 in the Asian Disease problem could have interpreted "combat the disease" in a number of different ways after reading the problem description. The problem setup though is interpreted by participants in the context of the two programs presented to each group. So let's examine some plausible interpretations of "combat the disease" Group 1 and Group 2 participants may have made either consciously or unconsciously given the specific context of each of their two programs.

Group 1 is presented specifically with Program A and Program B. Notice, however, that the consequences of Programs A and Program B are presented in terms of people who "will be saved".

[98] It is also possible that a combination approach is involved in the "combat the disease" programs involving perhaps a combination of prevention, treatment, and cure approaches and not just one program alone. This is actually more typical of real life disease fighting approaches as we have seen in the combating HIV disease example where a multi-faceted approach to disease fighting is seen as the most effective. If we were to assume that participants were interpreting "combat the disease" as being a combination of disease combating approaches this could only muddy the water further in case equalization.

Group 1's Program Options

If Program A is adopted, 200 people will be saved.
If Program B is adopted, there is 1/3 probability that 600 people will be saved, and 2/3 probability that no people will be saved.

A natural interpretation of "combat the disease" in a way that saves lives would be a disease fighting program that saves lives from disease infection. We know of cases where lives have been saved from disease infection by vaccines such as with the Polio vaccine. Participants might then naturally interpret "combat the disease" as vaccination programs that would save lives by preventing disease infection. Program A, for instance, could involve a known vaccine that is in limited supply such that only 200 people "will be saved" by the vaccine. Program B, in contrast, could perhaps be an experimental vaccine in ample supply but since the vaccine is experimental it may or may not be successful in preventing disease infection and saving lives. So Program B's experimental vaccine could have the scientific estimates of a 2/3 probability that no one will be saved and a 1/3 chance of 600 being saved.

Group 1 Vaccine Interpretation of Asian Disease Problem

Description:
Imagine that the U.S. is preparing for the outbreak of an unusual Asian disease, which is expected to kill six hundred people. Two alternative programs to combat the disease have been proposed. Assume that the exact scientific estimate of the consequences of the programs are as follows:

Group 1
Program A –Known Vaccine in Limited Supply; 200 people will be saved
Program B – Experimental Vaccine in Ample Supply; 1/3 probability that 600 people will be saved, and 2/3 probability that no people will be saved.

Observe that if participants in Group 1 were imagining this Known versus Experimental Vaccine interpretation, they would still most likely choose Program A over Program B just as Kahneman and Tversky found in their research. Participants in Group 1 could still view 200 people saved in Program A as their reference point and see Program B as a risky prospect since it could result in none of the 600 people being saved. On the interpretation of "combat the disease" which we are supposing, looking at the case with these more spelled out means arguably makes the moral reasons for the decision of Group 1 actually clearer. Group 1 believes 200 people can definitely be saved by the Known Vaccine even though there is a limited supply. Participants may believe it is morally wrong to choose Program B because it seems as if you are intentionally putting 200 lives at risk of disease and death who would not risk disease or death but for your decision. In other words, participants may be selecting Program A *only in part* due to the framing but may also be primarily

motivated by seeing Program B as involving Harmful Positive Agency or something similar to it since it is putting 200 lives at risk for a 2/3 chance risk of dying, that could otherwise have been saved for sure. Harmful Positive Agency recall is harm occuring because of what an agent does. We've seen that the choice of Program B puts 200 people at risk unnecessarily. By contrast, Program A's harms to the 400, who are not saved by the Known Vaccine, occur because of what the implementers of the program did not do, they chose to give the vaccine to some people (the 200) but not others (the remaining 400), and this is only Harmful Negative Agency. Program A involves, by contrast with choosing Program B, a simple traditional Doing and Allowing distinction, you allow 400 to die but it's the disease that kills them and not any Harmful Positive Agency on your part. This is just like our earlier example of a traditional case of Doing and Allowing in Foot's Rescue Dilemma 1, when we save the five from drowing but leave the one to drown because we cannot do both.

If this is the case, then the reason for participants in Group 1 choosing Program A over Program B could be due to some implicit acceptance of the moral distinction in the Doctrine of Doing and Allowing.

This interpretation for Group 1 can be contrasted with a possible interpretation of the Asian Disease Problem for participants in Group 2. Group 2 is presented with Programs C and D. Recall from our earlier presentation of the Asian Disease Problem that Programs C and D share the expected utility outcomes of Program A and B. Yet there is a difference in how Programs C and D are framed for

participants. Instead of being presented as "will be saved" as was the case for participants in Group 1, Group 2 has its options given in terms of "will die".

Group 2's Program Options

If Program C is adopted 400 people will die.
If Program D is adopted there is 1/3 probability that nobody will die, and 2/3 probability that 600 people will die.

Here a natural interpretation of "combat the disease" in a way that people "will die" is that the 600 people are already infected by the disease. We are also generally familiar with disease combating programs that help to prevent death for people infected with life threatening diseases. For instance, our earlier HIV antiretroviral medication example is a treatment that can enable people infected with the disease HIV to live. A Group 2 participant may naturally reason then that their "combat the disease" programs may be two different treatments for the infection that vary in effectiveness. Program C could be a Known Treatment that may not be effective for everyone infected by the disease.[99] Antiretroviral medication is an example of a kind of treatment that is not effective for everyone and can even stop working for some on the treatment.[100] The scientific estimates for Program C that

[99] We will try and make our Programs C and D somewhat equalized to Programs A and B but account for the vaguenesses. So we will use a Known Treatment for Program C versus an Experimental Treatment in Program D to make the programs more similar to Program A's Known Vaccine and Experimental Vaccine.
[100] Treatment 2011.

"400 people will die" could be considered a result of the Known Treatment only being expected to work for 200 of the 600 people.

Program D, on the other hand, could naturally be interpreted by Group 2 to be a risky Experimental Treatment for those infected by the disease. For instance, we are generally aware of terminally-ill cancer patients who may choose to take an extremely risky experimental treatment currently being tested by the FDA in the hopes that the treatment may save their lives.[101] Similarly, Program D is risky and may or may not prove effective. If Program D's is interpreted as an Experimental Treatment, then if the treatment works it will be effective for everyone but if the Experimental Treatment does not work everyone who takes the treatment will die. This can be drawn from the expected scientific estimates of Program D which specify that there is a "1/3 probability that nobody will die" but also a "2/3 probability that 600 will die".

[101] Joelving 2010.

Group 2 Treatment Interpretation of Asian Disease Problem

Description:
Imagine that the United States is preparing for an outbreak of an unusual Asian disease which is expected to kill six hundred people. Two alternative programs to combat the disease, A[C] and B [D], have been proposed. Assume that the exact scientific estimates of the consequences of the programs are as follows:

Group 2
Program C – Known Treatment that is not effective for everyone; 400 people will die.
Program D – Experimental Treatment that could be effective for everyone; 1/3 probability that nobody will die, and 2/3 probability that 600 will die.

With this treatment interpretation of "combat the disease" for Group 2, participants could reasonably select Program D over Program C which would be a decision that is consistent with Kahneman and Tversky's findings. Participants in Group 2 may be more willing to choose Program D and risk giving an experimental treatment to everyone infected by the disease, knowing that Program D would at least give everyone at risk of death a chance at living. The important thing here is that because of the "death" language the risk might be naturally conceived of as already involving the harm of infection. Whereas in Group I, it might be thought that everyone is fine, so far unharmed, and only being saved from possible harm. In Group 2 we are imagining that there is harm already, ie. people have contracted the disease, and we are only waiting for them all to die. A participant imagining things this way might think it is best to give everyone the same chances to survive the harm.

“Death” language vs. “saved” language is a framing that can have effects like these (as Prospect Theory constantly shows). The participants could be operating with a non-consequentalist intuition too, e.g., their choice of D could be based on the idea that it is “only fair,” or that it is what people would likely vote for, or some such. And depending on the details, some appeal to Doing and Allowing or Double Effect could enter in.

It's important to remember, too, that we need not (as Prospect Theory does not) assume that all of these imaginings or interpretations of the case are consciously thought out. Nor do we need to hold, as we have emphasized before, that such non-consequentialist intuitions could be defended to the hilt. Just that they would seem plausible to the participants in the study.

In this way Program D might seem more desirable in comparison to Program C where it is certain that 400 people will die. Prospect Theory indicates that Group 2 participants set their reference point at 400 people dying and this could still be the case for our Group 2 participants with a Treatment interpretation of the Asian Disease Case. However, the reasoning again, just as for our Group 1 Vaccine Interpretation, would not be sufficiently explained by appealing to the Prospect Theory's reference point alone. If Group 2 sees everyone as already being infected by the disease and 400 people “will die” as their reference point, the reason perhaps for choosing Program D is that by choosing Program D they give people who are already dying a chance at life. With the Experimental Treatment in Program D, we can foresee both

the possibility of all 600 people dying and the possibility of all 600 people living. Even if all 600 people die after being given the Experimental Treatment, we would not be directly intending their deaths only obliquely intending their deaths since it is only a known possible outcome. In this way, as noted, the non-consequentialist principle of Double Effect could be operating too.

Notice that all of these interpretations of "combat the disease" in the Asian Disease Problem bear on the question of the means of disease combating. Variation in the types of disease fighting programs and the targets of the programs show how interpretations of the means of "combat the disease" are variable. Furthermore, it is easy to see how combinations of the type of disease fighting programs with their respective disease combating targets offer other possible interpretations of what "combat the disease" means. We will discuss more of these later on but for now it is worth mentioning some of these possibilities. Preventing disease infections, for instance, could be prevention from a vaccine, prevention through education about transmission of disease, or through quarantine of those infected and numerous other possibilities.

Kahneman and Tversky seem to have to say that none of these possible interpretations matter in order for participants in Group 1 and Group 2 to come to the conclusions upon which they base their program choice. Prospect theory feels it can ignore these moral distinctions when considering how and why people choose the programs they do. This just shows how Prospect Theory is truly a consequentialist

theory as others, such as Francis Kamm, have pointed out.[102] At first it may have seemed that because Prospect Theory explains decision making better than Expected Utility Theory that the theory might be non-consequentialist in nature. However, as examining the Asian Disease Problem shows, Prospect Theory does not care about the means to the consequences. Prospect Theory only aims to be able to describe decision-making on the basis of the consequences of the action being considered. The reference point only sets up the manner in which the consequences are evaluated but does not change that the reason for the decision, according to Prospect Theory, is based upon the perceived consequences even if viewed in relation to a particular framing of the consequences.

This is especially clear in how Kahneman, even in recent discussions of the Asian Disease Problem, fails to see the unequalization in the case and continues to ignore the means. For instance, in his recent book "Thinking, Fast and Slow" published in October 2011, he uses the Asian Disease Problem as a central example for framing effects. It comes out too in how Kahneman mentions in the text how the Asian Disease Problem is just the same sort of problem as the Richer Decision Case. However, as we have seen, the Richer Decision Case is more clearly a problem dealing simply with maximizing money (see chapter 2, section 2.2, pages 12-27) whereas the Asian Disease Problem has more complex moral issues at stake.

[102] Kamm 2008.

Prospect theory, however, cannot ignore the variation in non morally equivalent means. For non-consequentialists, as we have seen, the means are critical to the moral permissibility of attaining certain ends. The Asian Disease Problem could be a case that shows specifically how ignoring the variation in non morally equivalent means may impact the results of studies even when the goal of the study is to assess preference for consequences. At the very least, however, the Asian Disease Problem does not provide clear evidence in support of Prospect Theory.

4.5 Additional Evidence of Equalization Problems

The Asian Disease Problem, however, is not the only instance where there is an issue with equalization and non morally equivalent means. Equalization has been known generally to be a problem for Kahneman and Tversky's case studies for Prospect Theory. Francis Kamm discusses problems with equalizing many of their other cases including the famous and still cited Snow Shovel Case.[103]

Just as we have seen in the Asian Disease Problem, Kamm argues that in the Snow Shovel cases Kahneman and Tversky have failed to ensure that all of the factors, other than the ones they are trying to test, are equal. In the Snow Shovel Loss Case a spring blizzard leads a store to raise the price on its snow shovels. However, in the Snow Shovel No-Gain Case a store does not reduce the price of its snow

[103] Kamm 2008.

shovels when it gets them cheaper from the dealer.[104] The Snow Shovel Loss Case is seen as unfair by participants studied, whereas, the Snow Shovel No-Gain Case is considered fair. In the Snow Shovel Loss Case, Kahneman argues, people believe they are entitled to the status quo snow shovel prices. So when the store raises the price on its snow shovels, due to the spring blizzard, Kahneman believes this normativizing (sense of entitlement) is the basis for the perceived unfairness. By contast, the Snow Shovel No-Gain Case is considered fair according to Kahneman since the customer is not entitled to a reduction from the status quo when the dealer is able to buy the snow shovels at a reduced cost (since the customer only normativizes the status quo snow shovel price). In other words, the customer expects the current price for snow shovels and does not think it is unfair if there is no price reduction since a price break on snow shovels would then be below his status quo expectations.

[104]According to Kahneman, due to a tendency to "normativize the status quo as something to which they have a right and are entitled, unless there are specific barriers to doing this (e.g. if the baseline is a sale, people would not think that they had a right to sale prices)." Kamm 2008, p. 428.

Snow Shovel Loss Case:

A spring blizzard leads a store to raise the price on its snow shovels.

Snow Shovel No-Gain Case:

A store does not reduce the price of its snow shovels when it gets them cheaper from the dealer.

Kamm points out that the Snow Shovel Loss Case and the Snow Shovel No-Gain Case have some factors which are not equalized. In the Snow Shovel Loss Case, there is a spring blizzard, whereas in the Snow Shovel No-Gain Case, there is no blizzard mentioned. Kamm points out that if we remove the spring blizzard from the Snow Shovel Loss Case, that the case no longer seems unfair (at least not obviously). This shows not only that the two cases are not equalized but also a problem with Kahneman's concept of the normativizing of the status quo. If there is a normativizing of the status quo, according to Kamm, we would still see the Snow Shovel Loss Case as unfair even without the blizzard. It should be considered unfair because the cost of the snow shovels would still be above the status quo expected price. Kamm also points out that what seems to be important is actually whether "the seller provides himself with a benefit relative to his status quo or instead prevents a loss to himself by raising the price, and whether he avoids imposing a loss on himself

when he does not lower the price."[105] This of course is a question with regard to the means in the case. The loss/no-gain distinction of Prospect Theory focuses on the outcomes for the customer, however, and not on the dealers themselves.[106] That is it focuses only on the victim, and what happens to him or her, rather than on the agency involved in making it happen. This again confirms the consequentialist strain in Prospect Theory.

I have dealt primarily with the Asian Disease Problem because it is central to Horowitz's argument, it is still very current, and most importantly because of how the Asian Disease Problem is connected to the Doctrine of Doing and Allowing and the Doctrine of Double Effect. The Asian Disease Problem also shows how broadly the means of the programs could be interpreted due to non equalization created by vaguenesses. What this indicates is that experimental evidence from the Asian Disease Problem does not clearly support Kahneman and Tversky's Prospect Theory.

So far we have looked at how a participant in Group 1 and a participant in Group 2 may have naturally interpreted the Asian Disease Case given the vaguenesses present in both the description and the programs. However, this is not the full extent of possible interpretations of either the description or the programs. While it is not possible to cover all of the possiblities, the general breadth of the vaguenesses can be laid out.

---

[105] Kamm 2008, p. 429.

[106] For non-consequentialists both the customer and the dealer are morally significant.

In general then it seems there are at least three categories of possible "combat the disease" interpretations: treatment, cure, or prevention.

Categories of "Combat the disease" Interpretations

Treatment: Combat the disease with a treatment for those infected.

Cure: Combat the disease with a cure to the disease.

Prevention: Combat the disease by preventing disease infections.

There are, however, other categorizations that are relevant to "combat the disease." Observe that some of these interpretations involve people who are already infected and some involve people who are not infected at all by the disease. A cure or treatment to the disease mainly seems to address those infected by the disease. In contrast, prevention would be more likely to be directed towards people who have not yet been infected by the disease in hopes of preventing infection.

However, it is worth noting that since we know nothing about the disease itself the aim of the "combat the disease" program does not have to be directed primarily at people at all. The disease could be transmitted via non-human carriers such as is the case with malaria which is mosquito-borne. A "combat the disease" program could be directed then at infected or not infected non-human carriers of the disease. For instance with malaria, a prevention disease combating program could use mosquito nets or insect repellants to prevent mosquito bites and thereby reduce disease transmission by infected mosquitos. There could also be direct intervention

aimed at the disease carrier. For instance with mosquitos, insecticides could be used to kill mosquitos that would harbor the disease. In general, one could argue, that it is reasonable to assume that particpants in the Asian Disease case were most likely thinking of programs directed at people rather than a non-human animal such as mosquitos for instance.

The example of mosquito-borne malaria, however, also demonstrates how disease combating programs could be interpreted as aiming at affecting the habitat or environment which helps the disease to flourish. This can be accomplished by disabling or interfering with transmission of the disease by a carrier, as is the case with mosquito nets and insect repellent which reduce mosquito bites which transmit the disease to their human hosts, or by affecting the habitat of the carrier, such as by draining standing water that is necessary for mosquito reproduction. It is also possible to direct a "combat the disease" program by targeting the immediate cause of the disease. Cures for instance could be given to a disease carrier but directly target the disease-causing bacterium. Penicillin, our earlier example, directly attacks the bacterium causing the disease, by affecting the bacterium's ability to construct its cell walls, resulting in the destruction of the bacterial cell.[107]

After reflecting on these various "combat the disease" categories, we can see several possible interpretations of disease combating targets, including programs that could be targeted at: infected people, not infected people, infected non-human carriers

[107] Shockman 1979.

of the disease, not infected non-human carriers of the disease, habitats or environments required for disease flourishing, and, finally, the immediate cause.

<u>"Combat the disease" Program Targets:</u>
Infected people
Not infected people
Infected Non-human carriers of the disease
Not infected Non-human carriers of the disease
Habitats or environments required for disease flourishing
The immediate cause, ie. a disease-causing bacterium

Reflecting on some of these possible interpretations of "combat the disease" meanings and their means to "combat the disease," it is clear that there are morally significant differences since not all of the various means are morally equivalent. For instance, the means used to "combat the disease" could be either involuntary or voluntary for people targeted by the programs. Voluntary participation of a person is frequently viewed in moral theory as being morally preferable to involuntary participation. The idea being, at least for nonconsequentialists, that voluntary involvement respects individual autonomy, his ability and right to make rational choices for himself, but that involuntary involvement in general violates a person's

autonomy.[108] Involuntary participation in a program could even be viewed as morally impermissible. Preventing disease infection by involuntary quarantine of infected people, for example, is not morally the same as preventing disease infection by educating those who are not infected by the disease about disease transmission. Involuntary quarantine of infected people impacts individual autonomy and rights. Education of those not infected by the disease about transmission by giving out educational pamphlets, for instance, does not usually involve similar harms to autonomy since reading the pamphlets is a voluntary act.[109]

There could be even more alternative interpretations with other significant moral issues. The "combat the disease" program could even be a treatment or program that requires allowing some innocent people to die to save the most amount of lives perhaps say by a deadly experimental treatment. An experimental disease combating treatment could be expected to kill people as a part of the treatment. A certain percentage of people would be expected to die from the treatment and perhaps

---

[108] Consequentialists, such as Utilitarians, may also find voluntary participation to be morally preferable to involuntary participation. A Utilitarian could argue that respect for individual autonomy will lead, in general, to the greater good. However, this is not necessarily the case since a Utilitarian could more easily justify acts such as involuntary quarantine if doing so was believed to result in creating, on balance, more good consequences. What is important here, however, is that even if a particular theory may or may not justify involuntary acts towards persons, most moral theorists would agree that the fact that an act was voluntary or involuntary is, on the whole, a morally significant difference.

[109] Voluntary action does not usually involve similar harms but could. Brainwashing, for instance, could be considered a case where voluntary action could still result in similar harms to autonomy. However, here I am simply demonstrating the idea that the differences between the possible various means of the disease-combating programs can be morally significant.

as many as who would die without any treatment. Cancer treatment and early small pox inoculation are such examples where the options involve life-threatening risks.[110]

Some "Combat the Disease" Means Interpretations:[111]

Treatment for those infected by the disease, eg. medicine that will result in fewer deaths but does not eliminate disease itself.

Treatment for those not infected by the disease, eg. vaccine.

Cure for those infected by the disease that eliminates the cause of the disease, eg. penicillin.

Preventing the spread of disease by prevention directed at those infected, eg. quarantine of those infected.

Preventing the spread of disease by prevention directed at those not infected, eg. education of people not infected on ways disease is transmitted.

Preventing the spread of the disease by targeting the disease carrier itself or environment, eg. for malaria one could drain still water habitats for mosquitos, or, directly kill the disease carriers such as killing mosquitos with insecticide.

---

[110] In particular I am thinking of chemotherapy for cancer patients where chemotherapy treatment itself could kill the cancer patient instead of the life-threatening cancer. Lansdown 2008. Early small-pox inoculations, also called variolation, involved exposing a healthy nonimmune person to the live small-pox virus. This exposure to the live small-pox virus was known to risk the death of the inoculated person from small-pox (2% to 3% died of the disease) but with 10 times lower fatality rates than with a typical exposure to small-pox. Riedel 2005.

[111] This list is not meant to be exhaustive but to be sufficient to illustrate some of the extent of the possible interpretations of "combat the disease."

The difference in each of these possible interpretations could be the same statistically in terms of outcomes of the programs but it is certainly not the same morally and shows even further the lack of equalization in the Asian Disease Problem. Furthermore, having different interpretations could affect the setting of the neutral reference outcome for each group, but more on this later. Before going into more detail about the problems raised by the vagueness of "combat the disease" in the description, we should turn to the vagueness in the phrasing of the programs' consequences.

In the two programs presented to Group 1 and Group 2 there are some language differences introduced that added more vagueness to the Asian Disease Case. Alternate phrasing of the options in general is intentional and is meant to elicit different subject responses. For example, if you tell a subject a product is 90% fat free, a participant may be more likely to say that the product does not have much fat in it then if you told them that the same product is 10% fat and then asked them if they thought the product had much fat in it. There is the same amount of fat in the product regardless of the phrasing of the facts. However, presenting either the phrasing of "fat free" or "fat" with the associated percentages on a product can elicit different participant responses even though nothing factual has changed about the

product itself.[112] The idea again being that people will make different program choices based more on how the programs are presented to the subjects and not on the basis of the statistical data that is identical in both groups' programs.

However, the alternate program phrasing in the Asian Disease case changes the nature of each program option to the point that drawing conclusions from comparisons between the Group 1 and Group 2 is problematic because the program options themselves have not been equalized.[113] Here again, just as we have seen with the vagueness of "combat the disease" in the description, the alternate phrasing of the two program options for Group 1 and Group 2 have vaguenesses that can create a significant difference which prevents sufficient equalization. Let us examine further how the phrasing in the programs for each group is vague.

Group 1's programs' consequences are phrased as people who "will be saved". In contrast, Group 2's programs's consequences are phrased as people who "will die." The phrasing of Group 1's programs as "will be saved" versus Group 2's programs as "will die" could not just create contrasting neutral reference outcomes in the same problem; the phrasing could also create an additional vagueness for the *means* of the programs being considered in each group. Both of the phrases also add a moral value to each of the programs' means, with one as morally positive and the

---

[112] Kahneman mentions this example in his recent book *Thinking Fast Slow*. See Kahneman 2011, p. 88.

[113] This is similar to what can be seen in the Snow Shovel Cases analyzed by Kamm discussed earlier. Kamm 2008.

other as morally negative. Part of the reason for this is due to vague implications in the use of the word "saved."

"Will be saved" implies both that the people are alive but also adds an additional dimension that the people are alive because of some positive action done in the means of the program that resulted in the lives being "saved." When we think of 'saving,' there is an implication that we are saving someone 'from' something. In the Asian Disease Case people "will be saved" perhaps from the disease itself, preventing infection and the subsequent death that could result, but people could also be saved from death alone and still be infected by the disease. Furthermore, being saved also implies something good was done for the person. In contrast, deciding that a group of people is "will die" in Group 2's program options does not have that positive inference and is construed as the opposite, as negative, since death has a generally negative connotation even more so for a certain death since they "will die." It is this question of the means of saving people that creates the main vagueness in the phrasing of the Asian Disease case programs.

The problem for Group 1 and Group 2 is still identical numerically but the phrasing of the programs imply different problems. Group 1 is being asked *how many lives to save*, ie. whether they would rather save 200 people or risk saving no one for a chance at saving 600. Group 2, on the other hand, is being asked *how many people can die*, ie. whether they would rather 400 die for certain, or risk 600 dying for a chance at no one dying. The alternate program phrasing for Groups 1 and 2 inclines

participants either consciously or unconsciously to perceive two distinct moral problems: how Group 1 views "saving" a life, versus, how Group 2 views choosing amount of people who "will die." At the very least it places Group 1 in a morally positive role of "saving" and Group 2 in a morally negative role of deciding how many are certain to "die." Using the phrase "will be saved" then indicates to Group 1 participants not just the program's consequences but also points to a positive means that brought about those consequences. The "saved" program has more positive means for Group 1 participants whereas the "will die" has more negative connotations for Group 2 participants.

This is quite different from how the contrasting neutral reference outcomes are elicited in the Richer Decision problem where participants were told at the onset they had different amounts of money. You could say having more money is better but this still does not seem to have the moral weight and value embedded in the choices themselves as they are in this Asian Disease case.[114] The problem could have, however, been phrased differently to perhaps avoid this problem by rewording Group 1's program in terms of "alive" or "will live" instead of "saved." There would still be the positive value of " will live" versus "will die" but these are more morally congruent than "saved" and "will die." We will look at this possibility more later but

[114] This is contrary to what Kahneman himself argues in *Thinking, Fast and Slow*. Kahneman believes the Richer Decision Problem and the Asian Disease Problem do not have such moral distinctions and should be viewed similarly. Kahneman 2011, p. 369.

for now it is sufficient to show that there could be phrasing that would not have been as morally loaded and vague.

Recall, the disease description states you are selecting disease combating programs but it is left open for interpretation what the means are, in other words, 'how' the programs bring about the consequences indicated. So participants in Group 1 are faced with not just one vaguenessof either "save" or "combat the disease," but two sets of vague wording viewed together in one problem. Taken up together the vagueness of "save" and "combat the disease" causes a combined vagueness creating a third unique larger problem of moral vagueness in the Asian Disease problem. This combined vagueness could make participants in Group 1 and Group 2, in other words, interpret different problems. Participants in Group 1, who have the "will be saved" programs, may be more inclined to interpret "combat the disease" as a positive life-saving action. Contrast this to participants in Group 2, who have the "will die" programs, who may be more prone to view "combat the disease" as a negative action that is more akin to choosing who dies. Our earlier examples of the Vaccine Interpretation for Group 1 and the Treatment Interpretation for Group 2 illustrate how all of these vaguenesses could come together specifically to exemplify not just the problem of nonequalization in general but also the nonequalization of means.

5. Summary and Concluding Remarks

Now that we have examined more fully the Asian Disease Problem, we can see that the Asian Disease Problem itself does not clearly support Prospect Theory since it was not properly equalized. On the contrary, the Asian Disease Problem, in at least our two plausible interpretations for Group 1 and Group 2, could actually be used to support both Quinn's view on the Doctrine of Doing and Allowing and the Doctrine of Double Effect. [115] At best, Prospect Theory could only explain our moral intuitions in part but cannot adequately capture them in full.

Prospect Theory, especially as a consequentialist moral theory, is unable to capture our intuitions by design as it is essentially unconcerned with the means that bring about the consequences. One could argue that Prospect Theory may be a better consequentialist theory than a standard Utilitarian account despite being more descriptive in nature. However, our main focus has been on whether or not there is a better explanation than the Doctrine of Doing and Allowing and the Doctrine of Double Effect for our moral intuitions. While it may be tempting for some to believe that there may be purely psychological underpinnings behind our decision-making, the Doctrine of Doing and Allowing is still the best explanation for our non-consequentialist intuions in moral dilemmas such as the Trolley Problem and others and it is not a psychological explanation but an explanation from moral theory.

[115] Kamm also has suspected this as well. Kamm 2008.

Prospect Theory clearly is limited in its ability to explain our decision making in rescue dilemmas like the Trolley Problem and even choices in the Asian Disease Problem originally used to support Prospect Theory. However, it is still not entirely clear what moral theorists should in the end take from Prospect Theory. The considerations here have led to issues in two areas: (1) the question of the interaction between psychological research and moral theory; and (2) the implications for a theory of moral decision-making that focuses on using the Doctrine of Doing and Allowing.

What is clear is that Prospect Theory provides at least an initial basis for critically looking at the interaction between psychological theory and moral decision-making. Our work here points to three main suggestions regarding the particular case study presented on the interaction between the Trolley Problem and other Rescue Dilemmas, the Doctrine of Doing and Allowing and Prospect Theory.

First, as many moral theorists may suspect, psychological research appears limited in how much it can explain issues of interest in moral theory. While it can be tempting to use psychology as a basis for moral positions, psychology is still descriptive in orientation versus prescriptive and it is this normative aspect of ethical decision-making that is missing necessarily from psychological research. This is demonstrated by what has been shown about Prospect Theory as it was originally posed by Kahneman and Tversky. Prospect Theory was specifically directed at addressing limitations of Expected Utility Theory as a behavioral model for decision-

making by creating a competing model. The normative role of evaluating what we 'ought' to do given the results of research still needs to be taken up by moral theorists. So in as much as psychological research does have implications to moral theory, as argued by philosophers such as Horowitz, this connection needs to be brought forward and applied into current and ongoing philosophical debates.

This presents the second main implication. Even though there may be limits to the type of moral explanations psychology can provide, the evidence given by psychological studies could be used to help support or discredit particular ethical theories. This is also evidenced by the example of Prospect Theory. As a model for decision-making, Kahneman and Tversky recognized some of the wider implications of Prospect Theory for moral value theory. So psychological research can have direct implications to ethical decision-making itself even if it does not directly tell us how we, as individuals or moral theorists, 'ought' to use this information.

Finally, given that there is applicable psychological research to moral theory, ethicists need to evaluate the results of research and integrate it into our moral arguments. As evidenced by the case study presented here on Prospect Theory, Doctrine of Doing and Allowing and the Trolley Problem, standing moral arguments can be further refined, challenged or strengthened by psychological theories. Philosophers such as Horowitz and Kamm have clearly taken seriously the implications of psychological research on moral theory.

Given these three implications of how we need to address psychological

research in moral theory, there is the further question: after we have examined how particular psychological theories are relevant to moral theory how do we proceed to develop the work in both fields? For the most part both disciplines continue on their separate paths and the interdisciplinary work of looking at the implications of work from one field on the other will have to be an intentional enterprise of individuals in both fields. Trying to show the importance of such continued interdisciplinary work is just part of the endeavor of this case study and dissertation.

I hope I have shown that the results of work in psychology continue to have relevance to moral theory and moral decision-making in particular, so far as certain experiments hold up and point to challenges to our more intuitive approaches in philosophy. One thing that seems clear is that Prospect Theory can help ethicists examine framing effects: it is worth understanding how our subjective frames of particular moral situations can influence what choices we feel are best. Moral decision-making theory needs to account for and explain results such as found in Prospect Theory by examining framing effects. In addition, as has been shown here, moral decision-making principles and cases, such as the Doctrine of Doing and Allowing and the Trolley Problem case in particular, have been critiqued using Prospect Theory framing effects as an alternative explanation. We may have adequately addressed initial concerns raised by the Asian Disease case about framing effects and moral theory. However, framing effects illustrated by Prospect Theory have presented moral decision-making theory an ongoing challenge to address

framing itself as a part of how to be better moral decision makers. And there of course can be refinements of the experiments, like the Asian Disease case, to pose further challenges. What has been shown in this thesis is why this particular experiment has fallen short of its goals: mainly because of the consequentialist strain in Prospect Theory (concern with the consequences at the expense of the means). This is a global issue with the theory, which might be overcom-able, but we have also seen how much attention would have to be paid to all the variations in the possible interpretations of key phrases in the framing of the cases meant to support the theory.

Appendix: Dilemmas and Cases

Asian Disease Problem

Description:

Imagine that the U.S. is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Assume that the exact scientific estimate of the consequences of the programs are as follows:

Group1

Program A; 200 people will be saved.

Program B; 1/3 probability that 600 people will be saved, and 2/3 probability that no people will be saved.

Group 2

Program C; 400 people will die.

Program D; 1/3 probability that nobody will die, and 2/3 probability that 600 will die.

Foot's Innocent Person and the Angry Mob

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

Foot's Rescue Dilemma I

…we are hurrying in our jeep to save some people – let there be five of them- who are imminently threatened by the ocean tide. We have not a moment to spare, so when we hear of a single person who also needs rescuing from some other disaster we say regretfully we cannot rescue him, but must leave him to die.

Foot's Rescue Dilemma I Rephrased

Six people are in danger of drowning but we can only either save one person from drowning or five from drowning (we cannot save everyone).

Foot's Rescue Dilemma II

…we are again hurrying to the place where the tide is coming in in order to rescue the party of [five] people, but this time it is relevant that the road is narrow and rocky. In this version the lone individual is trapped (do not ask me how) on the path. If we are to rescue the five we would have to drive over him. But can we do so? If we stop he will be all right eventually: he is in no danger unless from us. But of course all five of the others will be drowned.

Foot's Rescue Dilemma II Rephrased

You are traveling down a road to save five people from being killed. However, in order to save the five one must kill one person stuck in our path on the road.

Gambling Scenario

Option A: 50% chance to win $1,000, 50% chance to win nothing.

Option B: $450 for sure.

Horowitz's Rescue Dilemma I:

Option 1. - (a) Save a group of five people in danger of drowning, and (b) fail to save one person in danger of drowning.

Option 2. – (c) Save one person in danger of drowning, and (d) fail to save the group of five people in danger of drowning.

Horowitz's Rescue Dilemma II:

Option 3. – (e) Save a group of five people in danger of drowning, and (f) kill one person who would otherwise live.

Option 4. – (g) Fail to save the group of five people in danger of drowning, and (h) refrain from killing one person.

Marble Scenarios

Scenario 1: Close your eyes and draw a white marble from a clear vase that you can see in advance contains only one white marble.

Scenario 2: You see two marbles in a clear vase, one white marble and one black marble, and you blindly reach into the vase, you may still pull out a white marble but you will not know this for certain beforehand.

Scenario 3: Container A has two white marbles and one black marble, and Container B has two black marbles and one white marble. If you want to pull out a black marble and can only choose one container to draw from, then, like most people, you will choose to draw from Container B since you have only a 1/3 chance to draw a black marble from Container A but a 2/3 chance to draw a black marble from Container B.

Neighbor's Furnace Case

Background: Neighbor is old and might freeze if his furnace runs out of fuel and you have always helped the neighbor by firing up his furnace though you have not promised to do so.

Option 1: Fire up neighbor's furnace and do not help five equally close friends who need you to save them.

Option 2: Do not fire up neighbor's furnace, which leaves him in danger of freezing, in order to save five equally close friends.

Quinn's Rescue Dilemma III (a variant of Foot's Rescue II)

Scenario: You have taken over for the driver of a special train on a rescue mission to save five people when you discover someone trapped on tracks. The train is on automatic control but you can use the brakes to stop the train. If you allow the train to continue on the rescue mission to save five, the person stuck on the track will be killed by the train.

Option One: Stop train to save one person trapped on tracks, which aborts the rescue mission resulting in failure to save five.

Option Two: Let train automatically continue on tracks killing the one person trapped on the tracks in order to complete rescue mission to save the five people.

Quinn's Rescue IV (a variant of Foot's Rescue I)

Scenario: Five people injured on train that you are trying to save when you discover someone stuck on tracks. If you continue to save the five, the person stuck on the tracks will die.

Option One: Stop train to save one stuck on tracks and fail to save five injured on the train.

Option Two: Let train automatically continue on track and fail to save one in order to save five injured people on the train.

Richer Decision Problem

Group I
Setup: Participants told they are given $300.

Option One: Gain $100.

Option Two: 50% Gain Nothing, 50% Gain $200.

Group II
Setup: Participants told they are given $500.

Option One: Loose $100.

Option Two: 50% Loose Nothing, 50% Loose $200.

Snow Shovel Loss Case

A spring blizzard leads a store to raise the price on its snow shovels.

Snow Shovel No-Gain Case

A store does not reduce the price of its snow shovels when it gets them cheaper from the dealer.

The Trolley Problem

Imagine you are the driver of a trolley that is traveling out of control towards three innocent persons tied to the track. There is not time to remove these people from imminent danger, but you can steer the trolley away from them by pulling a lever and moving it onto an alternate track. However, this other track also has an innocent person tied to it who would be killed should the trolley be redirected.

Unemployed Gravedigger case

…some unemployed gravediggers, desparate for custom[ers], who got hold of this same oil and sold it (or perhaps *they* secretly gave it away) in order to create orders for graves. They strictly (directly) intend the deaths they cause, while the merchants could say that it was not part of their *plan* that anyone should die.

Wicked merchants case

In the wicked merchant case Foots says some "wicked merchants selling, for cooking, oil they knew to be poisonous and thereby killing a number of innocent people." The wicked merchants selling poisonous cooking oil can foresee the deaths of innocent people through the consumption of the

poisoned oil but they do not strictly intend for the people to die. The merchants are primarily intending to sell the poisonous cooking oil.

Appendix: Terms

Doctrine of Doing and Allowing: there are instances where it may be permissible to allow by inaction a harm to befall someone that one foresees will occur, even though one morally could not actively bring that same harm about by one's action.

Doctrine of Double Effect: the view that it is sometimes permissible to bring about by oblique intention what one may not directly intend.

Foot's Negative Agency: the allowing of a sequence of events to complete itself.

Foot's Positive Agency: to maintain or initiate a sequence that leads to harm.

Quinn's Harmful Negative Agency: harm occurring because of what the agent did not do but might have done (because of the noninstantiation of some kind of action he might have performed).

Quinn's Harmful Positive Agency: harm occurring because of what the agent does (because of the existence of one of his actions).

Bibliography

Appiah, Anthony. 2008. *Experiments in ethics*. Cambridge, Mass.: Harvard University Press.

Bentham, Jeremy. 1776. A Fragment on Government. London.

Broome, John. 1996. More Pain or Less? *Analysis* 56 (2):116-118.

Dawes, Robyn M., and Reid Hastie. 2001. *Rational Choice in an Uncertain World: The Psychology of Judgment and Decision Making*. Thousand Oaks: Sage Publications.

Doris, John and Stephen Stich. 2005. As a Matter of Fact: Empirical Perspectives on Ethics.

Driver, Julia. 2009. *The History of Utilitarianism*. The Stanford Encyclopedia of Philosophy 2009 [cited Summer 2009]. Available from http://plato.stanford.edu/archives/sum2009/entries/utilitarianism-history/.

Fischer, John Martin, and Mark Ravizza. 1992. Quinn on Doing and Allowing. *The Philosophical Review* 101 (2):343-352.

Foot, Philippa. 1967. *Theories of ethics*, *Oxford readings in philosophy*. London,: Oxford U.P.

———. 1984. Killing and Letting Die. In *Abortion: Moral and Legal Perspectives*, edited by J. L. Garfield and P. Hennesey. Amherst: The University of Massachusetts Press.

———. 2002a. *Moral dilemmas and other topics in moral philosophy*. Oxford and New York: Clarendon Press ; Oxford University Press.

———. 2002b. The Problem of Abortion and the Doctrine of the Double Effect. In *Virtues and vices and other essays in moral philosophy*. Oxford and New York: Clarendon Press ; Oxford University Press.

Foot, Philippa, Rosalind Hursthouse, Gavin Lawrence, and Warren Quinn. 1995. *Virtues and reasons : Philippa Foot and moral theory : essays in honour of Philippa Foot*. Oxford and New York: Clarendon Press ; Oxford University Press.

Fricker, Miranda. 1995. Intuition and Reason. *The Philosophical Quarterly* Vol. 45. (No. 179, April 1995):pp. 181-189.

Friedman, Alexander. 2002. Minimizing Harm: Three Problems in Moral Theory, Philosophy, MIT.

Garfield, Jay L., and Patricia Hennessey. 1984. *Abortion, moral and legal perspectives*. Amherst: University of Massachusetts Press.

Gonzalez, Richard, George Wu, and Jiao Zhang. 2004. Decision Under Risk. In *Blackwell Handbook of Judgment & Decision Making, Handbooks of Experimental Psychology*, edited by D. J. K. a. N. Harvey. Malden, MA: Blackwell Publishing.

Greene, Joshua. 2002. The Terrible, Horrible, No Good, Very Bad Truth about Morality and What to do about it, Philosophy, Princeton University.

Greene, Joshua and Haidt, Jonathan. 2002. How (and where) does Moral Judgment work? *TRENDS in Cognitive Sciences* 6 (12).

Hansson, Sven Ove. 2011. *Risk* 2011 [cited Fall 2011]. Available from http://plato.stanford.edu/archives/fall2011/entries/risk/.

Horowitz, Tamara. 1998. Philosophical Intuitions and Psychological Theory. *Ethics* 108 (2):367-385.

Howard-Snyder, Frances. 2011. *Doing vs. Allowing Harm*. The Stanford Encyclopedia of Philosophy 2011 [cited Winter 2011]. Available from http://plato.stanford.edu/archives/win2011/entries/doing-allowing/.

Joelving, Frederik. *The Perils of Taking Experimental Cancer Drugs*. Reuters.com, October 26, 2010. Available from http://www.reuters.com/article/2010/10/26/us-perils-drugs-idUSTRE69P4ZH20026.

Kahneman, Daniel. 2011. *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux.

Kahneman, Daniel, and Amos Tversky. 1979. Prospect Theory: An Analysis of Decision under Risk. *Econometrica* 47 (2):263-292.

———. 1981. The Framing of Decisions and the Psychology of Choice. *Science* 211 (4481):453-458.

Kamm, F. M. 1998. Moral Intuitions, Cognitive Psychology, and the Harming-Versus-Not-Aiding Distinction. *Ethics* 108 (3):463-488.

———. 2008. *Intricate ethics : rights, responsibilities, and permissible harm*, *Oxford ethics series*. Oxford ; New York ;: Oxford University Press.

Kamm, Frances M., and John Harris. 2000. The Doctrine of Triple Effect and Why a Rational Agent Need Not Intend the Means to His End: I - Frances M. Kamm. *Proceedings of the Aristotelian Society, Supplementary Volumes* 74:21-39.

———. 2000. The Doctrine of Triple Effect and Why a Rational Agent Need Not Intend the Means to His End: II - John Harris. *Proceedings of the Aristotelian Society, Supplementary Volumes* 74:41-57.

Kamm, Frances Myrna. 1986. Harming, Not Aiding, and Positive Rights. *Philosophy and Public Affairs* 15 (1):3-32.

Lansdown, M, M Mason, D Mort, K Protopapa, and N Smith. 2008. Systemic Anti-Cancer Therapy: For better, for worse? : National Confidential Enquiry into Patient Outcome and Death.

Lanteri, Alessandro, Chiara Chelini, and Salvatore Rizzello. 2008. An Experimental Investigation of Emotions and Reasoning in the Trolley Problem. *Journal of Business Ethics* 83 (4):789-804.

Mongin, Phillippe. 2012. *Expected Utility Theory*. Edward Elgar 1997 [cited April 13 2012].

Morgenstern, O. and J. Von Neumann. 1947. *Theory of Games and Economic Behavior*. 2nd (first edition: 1944) ed. Princeton, NJ: Princeton University Press.

Naylor, Margery Bedford. 1988. The Moral of the Trolley Problem. *Philosophy and Phenomenological Research* 48 (4):711-722.

Nichols, Shaun, and Ron Mallon. 2006. Moral dilemmas and moral rules. *Cognition* 100 (3):530-542.

Nozick, Robert. 1993. *The nature of rationality*. Princeton, N.J.: Princeton University Press.

Quinn, Warren. 1993a. Actions, intentions, and consequences: The Doctrine of Doing and Allowing. In *Morality and action*, edited by P. Foot. Cambridge [England] ; New York, NY, USA: Cambridge University Press.

———. 1993b. Actions, intentions, and consequences: The Doctrine of Double Effect. In *Morality and action*, edited by P. Foot. Cambridge [England] ; New York, NY, USA: Cambridge University Press.

Rickless, Samuel C. 1997. The Doctrine of Doing and Allowing. *The Philosophical Review* 106 (4):555-575.

Riedel, Stefan. 2005. Edward Jenner and the history of smallpox and vaccination. *Baylor University Medical Center Proceedings* v.18 (1):21-25.

Shockman, G. D., L. Daneo-Moore, J. B. Cornett, and M. Mychajlonka. 1979. Does Penicillin Kill Bacteria? *Reviews of Infectious Diseases* 1 (5):787-796.

Sinnot-Armstrong, Walter. 2006. Moral Intuitionism Meets Empirical Psychology. In *MetaEthics After Moore*, edited by T. Horgan and M. Timmons. USA: Oxford University Press.

Sinnott-Armstrong, Walter. *Framing Moral Intuitions* 2008. Available from http://sites.duke.edu/wsa/papers/files/2011/05/wsa-framingmoralintuitions2008.pdf.

Sinnott–Armstrong, Walter. 2002. Moral Relativity and Intuitionism. *Noûs* 36:305-328.

Thomson, Judith Jarvis. 1971. A Defense of Abortion. *Philosophy and Public Affairs* 1 (1):47-66.

———. 1985. The Trolley Problem. *The Yale Law Journal* 94 (6):1395-1415.