

VARIATIONS ON SOME ROSSIAN THEMES

A Dissertation Presented

by

KRISTIAN OLSEN

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

DOCTOR OF PHILOSOPHY

September 2015

Philosophy

© Copyright by Kristian Olsen 2015

All Rights Reserved

VARIATIONS ON SOME ROSSIAN THEMES

A Dissertation Presented

by

KRISTIAN OLSEN

Approved as to style and content by:

Fred Feldman, Chair

Peter Graham, Member

Ernesto Garcia, Member

Lyn Frazier, Member

Joseph Levine, Department Head
Department of Philosophy

ACKNOWLEDGEMENTS

Many people have helped me bring this dissertation to its present state. First and foremost, thanks to my advisor and the director of this dissertation, Fred Feldman. He has patiently read and commented on every draft of every chapter this dissertation contains (and several others that didn't make the cut). Without his tireless and lightning-fast feedback, it's difficult to imagine what this dissertation would look like. (I fear, though, that despite his best efforts, some parts of this dissertation are still not written in English. Forgive me Fred.) More generally, Fred has kindly shepherded me through each step of the graduate program. In fact, Fred was the first faculty member from the department that I met. He likes to tell the story of that meeting: I offered him my hand and said, "Hi, I'm Kristian"; he promptly shook my hand and replied, "Nice to meet you, I'm Jewish." I am thankful that we met and have worked together for so long, and I know that our friendship and philosophical partnership will continue long after this dissertation has been defended.

Special thanks also goes to Pete Graham, who has given me insightful feedback on versions of each chapter of this dissertation. I am also indebted to him for many long conversations about my ideas as they've developed. I can remember several occasions where our half-hour meeting turned into two hours or more. In many ways, this dissertation was born in a seminar on Ross and Intuitionism that Pete taught a few years back.

Thanks also to Ernesto Garcia for many fruitful conversations about many parts of this dissertation. He has been especially helpful in those places in this work where I

discuss Kant. I routinely go to Ernesto when I have a question about Kant or Kantian views more broadly, and he patiently listens to my puzzlement and helps set me straight. Plus, since he seems to have the entire Library of Congress stored in his head, he has been very helpful in directing me to the right books to read.

I am also grateful to Lyn Frazier for agreeing to serve as the outside member of my dissertation committee and for meeting with me on several occasions to discuss my work.

There are also many current and former graduate students to thank. Thanks especially to Miles Tucker, Josh DiPaolo, Luis Oliveira, Casey Knight, Matt Gifford, Lisa Cassell, Bob Gruber, Ed Ferrier, Brandy Burfield, Pete Marchetto, Dan McGloin, Justin Dealy, Rachel Robison, Jon Rosen, and Meghan Henning. They have all either given me helpful comments on portions of this dissertation, or provided me with much needed friendly distractions from my dissertation writing. The “or” here is inclusive, by the way.

Some other UMass faculty members also deserve special thanks. Thanks in particular to those who’ve led the dissertation seminars that I’ve been a part of: Maya Eddon, Joe Levine, and Louise Antony. I’ve presented versions of many chapters of this dissertation in dissertation seminar, and my work has always greatly improved as a result. Thanks also to Chris Meacham for helping me with some of the more technical parts of this dissertation and for the hard work he’s done to help me, and other grad students, develop professionally. Many thanks also to Beth Grybko, than which a greater office administrator cannot be conceived. She has, on countless occasions, helped me navigate through the inner workings of UMass and has provided me with good company in the department during the summers when things are quiet.

I also owe a debt of gratitude to my parents, Deb Ost and Steve Olsen, for giving me much needed emotional support during graduate school and for never questioning my desire to spend so much of my life studying philosophy.

Finally, and most importantly, I'd like to thank my wife, Leah Campbell. She has been by my side for every stage of this journey: she was with me for all of the highs and lows of my time at UMass; she was with me at Yale Divinity School when I didn't even know yet that I wanted to go to grad school for philosophy; and she was with me as an undergrad at Luther College when I barely knew any philosophy at all. Her constant love and support has carried me here.

ABSTRACT

VARIATIONS ON SOME ROSSIAN THEMES

SEPTEMBER 2015

KRISTIAN OLSEN, B.A., LUTHER COLLEGE

M.A.R., YALE DIVINITY SCHOOL

Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Fred Feldman

In this dissertation, I develop and defend some of W. D. Ross's moral views. Ross's views, I argue, are often highly plausible, though it is also often the case that variations on (or modifications to) his views are needed in order to remain philosophically tenable. In my dissertation, I explain why these variations are necessary and what they should look like.

In chapter 1, I discuss Ross's theory of moral rightness in his most important work, *The Right and the Good*. In chapters 2 and 3, I correct various misunderstandings about Ross's position: I argue that he is no more a particularist about absolute duty than a utilitarian or a Kantian is, and on many definitions of "pluralism" present in the literature, he is not in fact a pluralist, as he is typically assumed to be. In chapter 4, I discuss several objections that Ross later comes to make to his own theory of rightness; I argue, however, that none of them are any good. In chapter 5, I argue against Ned Markosian's recent claim that "Rossian Minimalism" is the best theory of rightness that makes use of the concept of a *prima facie* duty: Ross's own theory (or, at least, a "world" version of it) is, I maintain, superior to Rossian Minimalism. In chapter 6, I address some objections to

Ross's theory suggested by Michael Stocker and Michael Slote and demonstrate that the best way of responding to them is by transforming Ross's theory into a "dual-ranking" one. In chapter 7, I discuss Ross's theory of the subjective sense of "right" (chapters 1-6 are primarily concerned with Ross's theory of the objective sense of the term). I show that Ross's theory is problematic, and I offer a better theory in its place. In chapter 8, I turn to Ross's theory of moral goodness. I argue that his theory is more plausible than other theories suggested in the literature, but it suffers from the "nepotism problem." I show that Ross's solution to this problem is unsatisfactory and suggest a better way forward.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	iv
ABSTRACT	vii
LIST OF TABLES	xii
INTRODUCTION	1
 CHAPTER	
1. ROSS’S THEORY OF RIGHTNESS IN <i>R&G</i>	5
1.1 Introduction	5
1.2. Historical Context	5
1.3. The Concept of a Prima Facie Duty	8
1.4. Ross’s Theory of Rightness	15
1.5. Our Prima Facie Duties, According to Ross	25
1.5.1. Fidelity	25
1.5.2. Gratitude	26
1.5.3. Reparation	26
1.5.4. Beneficence, Self-Improvement, and Justice	27
1.5.5. Non-Maleficence	32
1.5.6. Promise Breaking, Ingratitude, Unreparation, Injustice?	33
1.5.7. Bringing about Intrinsic Badness?	33
1.5.8. Summary	34
1.5.9. Loose Ends	34
1.6. Ross’s Theory Compared with Ideal Utilitarianism and Kantianism	39
2. ROSS AND PARTICULARISM	44
2.1. Introduction	44
2.2. Utilitarianism and Kantianism	45
2.3. What’s an Absolute Moral Principle?—Some Initial Answers	47
2.4. What’s an Absolute Moral Principle?—Another Answer	54
2.5. What’s an Absolute Moral Principle?—Some Final Answers	60

2.6. Conclusion and Application to Dancy	65
3. ROSS AND PLURALISM	71
3.1. Introduction.....	71
3.2. Pluralism(c).....	72
3.3. Pluralism(n&s).....	78
3.4. Pluralism(3).....	80
3.5. Pluralism(2).....	84
3.6. Conclusion	87
4. ROSS'S OBJECTIONS TO ROSS.....	88
4.1. Introduction.....	88
4.2. Ross's Theory of Moral Obligation in <i>FE</i>	88
4.3. Argument #1	97
4.4. Argument #2	101
4.5. Argument #3	106
4.6. Some Final Arguments	112
4.7. Conclusion	116
5. ROSSIAN MINIMALISM VS. WORLD ROSSIANISM	117
5.1. Introduction.....	117
5.2. Some Alternatives to Russian Minimalism.....	117
5.3. The Bergström/Castañeda Problem	121
5.4. World Rossianism.....	126
5.5. Problems for World Rossianism	131
5.6. Russian Minimalism	137
5.7. Other Problems for Russian Minimalism	144
5.8. Conclusion	149
6. THE PROBLEM OF SELF-BENEFIT AND SELF-HARM	150
6.1. Introduction.....	150
6.2. First Response: Embrace the Implications.....	152
6.3. Second Response: Appeal to Praiseworthiness and Blameworthiness	153
6.4. Third Response: Deny the Prima Facie Duties	155
6.5. Fourth Response: Dual-Ranking.....	157
6.6. Dual-Ranking, Take 2.....	163
6.7. Dual-Ranking, Take 3.....	166

6.8. Appendix: Combining Dual-Ranking Rossianism ³ and World Rossianism	173
7. ROSS'S THEORY OF SUBJECTIVE RIGHTNESS	175
7.1. Introduction.....	175
7.2. Some Alleged Problems for Ross's Theory, and Some Real Ones	178
7.3. Minimizing Expected Objective Wrongness	187
7.4. Smith and Feldman	193
7.5. SR _{MEOW} and Action-Guidance.....	202
7.6. A Problem Involving Supererogation	209
7.7. The Imprecise Justificatory State Problem	214
8. ROSS'S THEORY OF MORAL GOODNESS	221
8.1. Introduction.....	221
8.2. Ross's Theory (And Some Background)	222
8.3. The Coincident Reasons Thesis	231
8.4. The Nepotism Problem	236
8.5. The Solution.....	247
8.6. An Objection.....	257
8.7. Conclusion	263
8.8. Appendix: Markovits on Fundamental and Non-Instrumental Reasons.....	264
BIBLIOGRAPHY.....	275

LIST OF TABLES

Table	Page
1. Keeping a slightly sub-optimal promise	41
2. Breaking a promise to save a life	43
3. A problem for Bergström's conception of alternatives.....	125
4. Jolt of Pleasure	160
5. Jolt of Pain	161
6. Money	162
7. Appendages	163
8. World Poverty	164
9. World Hunger according to Dual-Ranking Rossianism ²	165
10. Fire	168
11. Fire, updated	169
12. Dr. Jill (expected utility)	185
13. Dr. Jill (expected objective wrongness).....	190
14. Fire (Second Version)	210
15. Fire (Third Version)	212
16. Fire (Third Version), updated	213

INTRODUCTION

W. D. Ross's most enduring philosophical work is devoted to two central topics in ethical theory: the right and the good (the title of the work in question is therefore unsurprising: *The Right and the Good*). In this dissertation, I discuss Ross's views about the right. I ignore almost entirely Ross's views about the good, though I do discuss in the final chapter his views about *moral* goodness (i.e., his views about moral worth, or moral praiseworthiness). My interest in Ross's views about the right (and about moral goodness) is not primarily historical, though I do think his views about these matters are interesting for historical reasons: Ross was an important figure in the school of British moral philosophy in the late 19th and early 20th century that spans roughly from Henry Sidgwick to A. C. Ewing.¹ However, I am not chiefly interested in Ross because he was an important historical figure in an influential school of ethics; rather, I am interested in Ross because his views, at least about the right and about moral goodness, seem to me to be plausible. The goals of this dissertation are (1) to explain, as clearly as possible, what these views are; and (2) to defend them, at least as much as possible. As we'll see, Ross's views are not entirely unproblematic. In order to make them plausible, they will often need some modifications (or *variations*—hence the title of this dissertation). However, I do think his views are often close to the truth. At the very least, they are excellent starting points to getting there. By the end of this dissertation, I hope the reader is convinced of this, even if she is not inclined to accept the particular variations of Ross's views that I leave her with.

¹ Sadly, this period in the history of ethics has no agreed-upon name, though perhaps “British Intuitionism” is an apt one. For a nice overview of this period, see Hurka (2011).

Here is a brief description of the chapters this dissertation contains.

In chapter 1, I discuss Ross's theory of the right (or, as I prefer to call it, his "theory of rightness") in *The Right and the Good*. I argue that in that work, Ross holds that an action is right if and only if (and because) its total prima facie rightness minus its total prima facie wrongness is at least as great as that of any other action the agent can perform instead. In the process of so arguing, I situate Ross's theory in its historical context, discuss his notion of a prima facie duty, and explain what prima facie duties he thinks we have.

Much of the rest of the dissertation deals with objections to Ross's theory of rightness from *The Right and the Good* (hereafter abbreviated as "*R&G*"). However, before discussing these objections, I devote two chapters to correcting various misunderstandings about Ross's theory. In chapter 2, I discuss the widespread belief that Ross is a particularist about absolute duty. I argue that this belief is misguided: Ross is no more a particularist about absolute duty than a utilitarian or a Kantian is. In chapter 3, I discuss the common claim that Ross is a pluralist. I argue that this claim is also mistaken: I show that on four common ways of understanding what pluralism is, Ross is not in fact a pluralist, or, at least, we have no good reason to think that he is.

In chapter 4, I turn to objections. Chapter 4 is devoted to objections that Ross eventually comes to make to his own theory of rightness from *R&G*. These objections appear in Ross's later work, *The Foundations of Ethics* (hereafter abbreviated as "*FE*"). In that work, Ross alters his theory of rightness significantly and argues that his earlier theory of rightness is mistaken. In chapter 4, I discuss the details of Ross's new view and

examine his arguments for adopting it over the view he holds in *R&G*. I argue, however, that these arguments are unpersuasive.

Chapter 5 is devoted to Ned Markosian's recent paper, "Rossian Minimalism." In that paper, Markosian argues that the most plausible theory of rightness that makes use of Ross's notion of a prima facie duty is a view he calls "Rossian Minimalism." According to Rossian Minimalism, an action is right if and only if (iff) it minimizes total prima facie wrongness. I argue, however, that Markosian is mistaken: Ross's own view (from *R&G*) is the most plausible theory of rightness that makes use of Ross's notion of a prima facie duty. I do concede, however, that to remain the best, some technical modifications to Ross's theory need to be made. In particular, Ross's theory needs to be made into a "world" theory. I discuss in detail what such a theory looks like.

In chapter 6, I consider an objection to Ross's theory that is inspired by the work of Michael Stocker and Michael Slote. Ross's theory, when combined with his claim (in *R&G*) that there is a prima facie duty to benefit oneself and a prima facie duty not to harm oneself, has implications that, Stocker and Slote would insist, are counterintuitive. However, I argue that it won't do for a Rossian to merely abandon these prima facie duties, for self-benefit and self-harm seem relevant to the determination of rightness even if we do not have the prima facie duties in question. In order to respond to these issues, a Rossian should, I maintain, transform her theory into a "dual-ranking" one (*a la* Doug Portmore) according to which the moral rightness of an action is a function of not only the agent's prima facie moral duties, but her prima facie prudential duties as well. In chapter 6, I explain what this function should look like.

Chapters 1-6 are primarily concerned with Ross's theory of the *objective* sense of "right." But Ross thinks that alongside an objective sense of "right," there is also a subjective sense of the term. The objective sense expresses what is right for us to do given the (mind-independent) facts of our situation; the subjective sense, on the other hand, expresses what is right for us to do given our beliefs about (or evidence concerning) those facts. I am in broad agreement with Ross that we need to distinguish between these two senses of "right." However, I think that Ross's theory of the subjective sense (which he offers in *FE*) is problematic. In chapter 7, I explain why, and I then offer what I think is a better theory. According to it, an action is right (in the subjective sense) iff it minimizes expected objective wrongness. I unpack this theory in chapter 7, compare it with some other theories suggested recently in the literature, and defend it from several objections.

In the final chapter of this dissertation, chapter 8, I switch gears a bit: I discuss Ross's theory of moral goodness. According to the theory he offers in *R&G*, an action is morally good iff it is done from a good motive. Furthermore, Ross thinks that there are several good motives. In chapter 8, I argue that Ross's view is highly plausible; in fact, it is more plausible than other theories of moral goodness on the market, including Kant's well-known theory and Julia Markovits's and Nomy Arpaly's Coincident Reasons Thesis. Ross's theory, however, is not problem-free: it is subject to a problem I call the "nepotism problem." Ross is aware of this problem, and he tries to solve it in *FE*. I argue, however, that Ross's solution is unsatisfactory, and I suggest a better solution in its place, a solution that posits the existence of multiple grounds of praiseworthiness and blameworthiness.

CHAPTER 1

ROSS'S THEORY OF RIGHTNESS IN *R&G*

1.1 Introduction

In this chapter, I will discuss what Ross's theory of rightness¹ is in *R&G*.² In the course of doing this, I will place Ross's theory in its historical context, discuss Ross's notion of a prima facie duty, and explain what prima facie duties he thinks we have.

1.2. Historical Context

In order to fully appreciate Ross's theory of rightness in *R&G*, it will be useful to briefly situate his theory in its historical context. Ross's theory is a reaction to both utilitarian and Kantian theories of rightness. Ross proposes his theory as a way of solving what he sees as decisive problems with these theories.

The most plausible version of utilitarianism, Ross thinks, is G. E. Moore's version of it (commonly referred to as "Ideal Utilitarianism"). However, as Ross recognizes (1930, 6-11), it's not entirely clear how Moore's theory should be understood. The version of utilitarianism that Moore seems to adopt in *Principia Ethica* ([1903] 1993) is a claim about *meaning*. According to it, the term "right" means "productive of the greatest possible good." However, as Ross (1930, 8) rightly points out, this view is implausible.

¹ More precisely, I will be discussing Ross's theory of *moral* rightness (as opposed to his theory of prudential rightness, legal rightness, etiquettical rightness, etc.). Unless otherwise noted, I will always have moral rightness in mind in this dissertation.

² In chapter 4, I will discuss what Ross's theory of rightness is in *FE*.

Indeed, it is subject to the same type of criticism that Moore uses to undermine the claim that the term “good” means “pleasant.” If “right” meant “productive of the greatest possible good,” then the sentence “right actions produce the greatest possible good” would be trivially true. But clearly it’s not. Additionally, Ross (1930, 9) suggests that if “right” meant “productive of the greatest possible good,” then whenever a person says that an action is right, he would need to be thinking about its total consequences. But ordinary people surely don’t think about this every time they assert that an action is right.

Perhaps for these reasons, Moore seems to reformulate his version of utilitarianism in his later book, *Ethics* ([1912] 2005). There, Ross (1930, 10-11) points out, Moore formulates utilitarianism not as a view about the meaning of “right,” but as a view about the *ground* of rightness. In other words, his version of utilitarianism is supposed to indicate not what the term “right” means, but rather, what *makes* right actions right (or what *explains why* right actions are right). His view can be stated as follows:

An action is right if and only if (and because) it would bring about a greater balance of intrinsic goodness over intrinsic badness than any of its alternatives would bring about.³

This view, Ross rightly suggests, is much more plausible than Moore’s earlier view, and it is to it that I will henceforth be referring when I use the term “Ideal Utilitarianism.”

While Moore’s Ideal Utilitarianism may be the most plausible version of utilitarianism, Ross (1930, 34-39) thinks it is deeply problematic. Suppose, Ross suggests, that I have two options: either I can (i) produce 1,000 units of intrinsic

³ Here, following Ross, I am using the term “right” to mean “obligatory” (instead of “permissible”). I will return to this point later.

goodness for person A by fulfilling a promise I have made to him, or (ii) produce 1,001 units of intrinsic goodness for person B by breaking my promise to A. Suppose that my options will produce no other good or bad consequences. Given that, Ideal Utilitarianism implies that it is right for me to perform option (ii) and not right for me to perform (i). According to Ross, this is deeply counterintuitive. He says:

We should, I fancy, hold that only a much greater disparity of value between the total consequences would justify us in failing to discharge our *prima facie* duty to A. After all, a promise is a promise, and is not to be treated so lightly as the theory we are examining [Ideal Utilitarianism] would imply. What, exactly, a promise is, is not so easy to determine, but we are surely agreed that it constitutes a serious moral limitation to our freedom of action. To produce the 1,001 units of good for B rather than fulfil our promise to A would be to take, not perhaps our duty as philanthropists too seriously, but certainly our duty as makers of promises too lightly. (1930, 35)⁴

At this juncture, it is useful to note that Kant's theory of rightness makes the correct conclusion about this case, for according to him, we have various "perfect" duties, such as the duty to tell the truth, to pay our debts, and to fulfill our promises (see Ross 1930, 18). Whatever else it might mean for a duty to be "perfect," a perfect duty is one that is never permissible to violate. So, if there is a perfect duty to keep our promises, as Kant seems to have thought, then it will follow that it is always wrong for us to break our promises. Kant's theory therefore implies that in the case presented above, I should perform option (i).

However, Kant's theory is no less problematic than Ideal Utilitarianism is. Quite simply, it is implausible to hold that we have the perfect duties that Kant says we have. Consider promise keeping, for instance. It is difficult to accept that we have a perfect

⁴ In *FE*, Ross makes several additional objections to Ideal Utilitarianism (see 1939, 67-79).

duty to keep our promises since there seem to be scenarios where it is clearly permissible to break them. As Ross says, there are “exceptional cases in which the consequences of fulfilling a promise (for instance) would be so disastrous to others that we judge it right not to do so. It must of course be admitted that such cases exist. If I have promised to meet a friend at a particular time for some trivial purpose, I should certainly think myself justified in breaking my engagement if by doing so I could prevent a serious accident or bring relief to the victims of one” (1930, 18; see also 28). Similar things can be said about lying and failing to repay debts. If a situation were to arise where we could save someone’s life by telling a trivial lie or by failing to repay a small debt, it seems permissible to do so. It is therefore difficult to maintain, as Kant does, that we have a perfect duty to tell the truth and to repay our debts.

It is in this context that Ross presents his theory of rightness. His theory, he thinks, evades the problems of both Ideal Utilitarianism and Kantianism. However, before we can see why this is the case—indeed, before I can even state Ross’s theory—I need to discuss the concept of a *prima facie* duty, for Ross’s theory of rightness crucially involves it. It is to this task that I now turn.

1.3. The Concept of a *Prima Facie* Duty

Ross provides no shortage of descriptions of *prima facie* duties. In *R&G*, he likens *prima facie* duties to conditional duties (1930, 19, 23), *ceteris paribus* duties (30), claims (20), tendencies (28), and natural laws (28-29). In *FE*, Ross additionally describes *prima facie* duties as responsibilities (1939, 85), intuitions (82-83), things fitting to the situation

(81-84), suitabilities (85, 315-17), arguments (106), and ethically relevant characteristics (86).

I suspect that many will find these descriptions less than fully satisfying. What Ross means by “tendency,” “parti-resultant attribute,” “suitability,” etc. seems at least as obscure as what he means by “prima facie duty.” Perhaps for these reasons, many philosophers now understand prima facie duties in a way not explicitly suggested by Ross. Many now take a prima facie duty to simply be a moral reason.⁵ More precisely, many would urge us to accept the following:

PFD: There is a prima facie duty for a person to perform an action A =_{df.} there is a moral reason for the person to perform A.

I agree; I think we should accept PFD. Although Ross never uses the term “moral reason” in his work—the term doesn’t seem to enter the philosophical lexicon until around 1950, well after Ross wrote his main philosophical works—there is good reason to identify Ross’s concept of a prima facie duty with the concept of a moral reason. After all, moral reasons seem to have many (perhaps all) of the features that Ross ascribes to prima facie duties. I will now enumerate several of these features:

1. In perhaps the most famous passage from *R&G*, Ross claims that there are prima facie duties of (1) fidelity (i.e., promise keeping), (2) reparation, (3) gratitude, (4) justice, (5) beneficence, (6) self-improvement, and (7) non-maleficence (1930, 21). I’ll discuss each of these claims in detail below. For now, I simply want to point out that these claims seem plausible if we take prima facie duties to be

⁵ See, e.g., Feldman (1978, 150-52), Dancy (1993, chap. 6; 2004b, 18-21), Stratton-Lake (2000, chap. 5; 2002, xxxiii-xxxviii; 2011a, 2011b), and Markosian (2009, 2).

moral reasons. In other words, it seems plausible to hold that there is a moral reason to keep one's promises, to make reparations for past wrongs, to show gratitude, to benefit others, etc.

2. Ross claims that prima facie duties have stringencies (1930, 21, 25, 41, 54, 55). He says that one prima facie duty can be more "incumbent" (19), "urgent" (23), or "pressing" (31) than another. Similarly, moral reasons also have strengths. If there is a moral reason to perform an action, it makes sense to ask how *strong* or *weighty* the moral reason is. Sometimes there is a strong moral reason to perform an action; other times, there is only a weak moral reason to perform an action.
3. Ross thinks there's an important difference between a prima facie duty and what he calls an "actual" (or "absolute") duty (1930, 20). (Philosophers now often use the terms "overall duty" and "all things considered duty" to refer to what Ross means by "absolute duty.") Prima facie duty and absolute duty are not the same thing, for there might be a prima facie duty to perform an action even though, overall, one ought not to perform it. According to Ross, this will happen when, for instance, keeping a promise will have disastrous consequences. In such a case, there is a prima facie duty for the promiser to keep the promise, but she shouldn't, all things considered, keep it. Moral reasons also have this feature. There can be a moral reason to do something even though it would be wrong, overall, to do it. Intuitively, this will happen in the promising case just described: there is a moral reason for the promiser to keep the promise, but she shouldn't keep it, all things considered.

4. Ross likens *prima facie* duties to forces (1930, 28-29). He notes that some of the individual forces acting on an object might be in a different direction from the direction the object is actually moving. This will happen when the forces exerted on the object in one direction are combined with forces exerted on the object in a different direction. *Prima facie* duties, Ross suggests, are like individual forces acting on an object; absolute duties are like the product of these individual forces (or, perhaps, the actual motion of the object). Ross notes, however, that this comparison isn't perfect, for *prima facie* duties are in no way *causal* forces. Still the metaphor is illustrative. The metaphor is also apt, I think, when it comes to describing moral reasons. While most philosophers these days refuse to analyze the concept of a moral reason, many will go so far as saying that for there to be a moral reason to do something is for there to be a moral consideration that counts in favor of it.⁶ But the notion of a consideration *counting in favor* of an action seems very much like the notion of a *force* acting on the action. Though again, the comparison isn't entirely unproblematic: moral reasons are surely not causal forces.
5. Ross is not happy with the term “*prima facie* duty” (1930, 20). For one thing, it suggests that a *prima facie* duty is a certain kind of *duty*, but according to Ross, a *prima facie* duty is no sort of duty at all. For another, the term suggests that a *prima facie* duty is something that is merely *apparent*, an illusion that vanishes after further inspection, but according to Ross, a *prima facie* duty is no such thing; a *prima facie* duty is “an objective fact involved in the nature of the situation”

⁶ See, e.g., Scanlon (1998, 17), Dancy (2004b, 29), and Parfit (2011, 31).

(20). Similar things can be said of moral reasons. A moral reason is not a kind of duty. That is, to say that there is a moral reason to do something is not to say that it ought to be done. Indeed, one presumably ought *not* to do it if there are stronger moral reasons against it. Also, when we say that there is a moral reason to do something, we are not saying something epistemological. We are not saying that it *appears* or *seems* that we should do it. Rather, we are picking out an objective feature of the action—we are saying that, as a matter of fact, there is a moral consideration that counts in favor of it.

6. Ross claims that when there is a prima facie duty to perform an action, the action would be an absolute duty if it had no other “morally significant” features (1930, 19, 138). This also seems true of moral reasons: if there is a moral reason to perform an action, and the action has no other features of moral relevance, then the agent of the action ought to perform it.⁷

So, I think that if we understand Ross’s concept of a prima facie duty in terms of the concept of a moral reason, we can make sense of much, if not all, of what Ross says about prima facie duties in the text. I thus propose that we understand prima facie duties in this way.

Before moving on, I’d like to make a few clarifying remarks about PFD from above. First, PFD provides a definition for the locution “there is a prima facie duty for a person to perform an action.” Ross sometimes uses this phrasing (see, e.g., 1930, 25, 58). However, he also uses other locutions that involve the term “prima facie duty.” He talks

⁷ Notice that I merely say that this *seems* true. In later chapters, we’ll see some reasons to doubt it (see, especially, chaps. 5-6).

of a person's "having" a prima facie duty to do something (see, e.g., 24, 35, 58, 61), and he talks of an action's "being" a prima facie duty (see, e.g., 28, 29, 30, 33, 35, 40, 46, 62). However, I will assume that for a person to *have* a prima facie duty to do something just is for there to be a prima facie duty for the person to do it. Likewise, I will assume that for an action to *be* a prima facie duty is for there to be a prima facie duty for the agent to perform it. In this way, we can use PFD to understand all of Ross's uses of "prima facie duty."

Second, in addition to talking of an action's being a "prima facie *duty*," Ross also frequently talks of an action's being "prima facie *right*" (see 1930, 29-47). However, Ross seems to use the term "prima facie right" to simply mean "prima facie duty." So, I will assume that when Ross speaks of an action's being "prima facie right," he means that the action is a prima facie duty (which, in turn, means that there is a prima facie duty for the agent to perform it).

Third, Ross not only thinks that there can be prima facie duties *to* do things, he also thinks that there can be prima facie duties *not* to do things (1930, 22, 26, 55, 61). Using PFD as our model, we can give a corresponding account of what it means to say that there is a prima facie duty not to do something:

~**PFD**: There is a prima facie duty for a person *not* to perform an action A =_{df.} there is a moral reason for the person *not* to perform A.

Ross also talks of a person's "having" a prima facie duty not to do something, and he talks of an action's "being" prima facie wrong. As before, we can understand these things simply to mean that there is a prima facie duty for the agent not to perform the action.

Fourth, PFD is simplified in the following way. Ross thinks that when there is a prima facie duty to perform an action, this is always in virtue of the fact that the action has some property or feature. For this reason, Ross calls the property of being a prima facie duty a “parti-resultant” property—i.e., it’s “one which belongs to an act in virtue of some one component in its nature” (1930, 28). (In contrast, the property of being an absolute duty is, Ross says, a “toti-resultant” property—i.e., it’s “one which belongs to an act in virtue of its whole nature and of nothing less than this” (28).⁸) More accurately, then, we should state PFD as follows:

⁸ Ross later acknowledges that this is an overstatement. He notes that “Any act is the origination of a great variety of things many of which make no difference to its rightness and or wrongness,” and the property of being an absolute duty does not belong to an action in virtue of these things (1930, 33n2). However, Ross does insist that “there are always many elements in its [an action’s nature] (i.e., in what it is the origination of) that make a difference to its rightness or wrongness, and no element in its nature can be dismissed without consideration as indifferent” (33n2). Ross, then, seems to be suggesting that while the property of being a prima facie duty belongs to an action in virtue of one feature that makes a difference to its rightness or wrongness, the property of being an absolute duty belongs to an act in virtue of all of these features, where the set of an action’s features that make a difference to its rightness or wrongness is a subset of the set of all its features.

However, this way of distinguishing between prima facie duty and absolute duty is confusing for the following reason. As we’ve seen, Ross thinks that an action is an absolute duty in virtue of having a greater total balance of prima facie rightness over wrongness than that of any alternative. So, it seems that Ross would have to admit that the property of being an absolute duty *is* in fact determined by one feature that makes a difference to its rightness or wrongness, just as the property of being a prima facie duty is.

How should Ross be interpreted here? Perhaps like this. Call a feature that makes an action a prima facie duty a “morally relevant feature.” Ross’s claim is that the property of being a prima facie duty belongs to an action in virtue of *one* of its morally relevant features. The property of being an absolute duty also belongs to an action in virtue of one of its features—namely, the feature of uniquely maximizing the total balance of prima facie rightness over wrongness. However, *this* feature belongs to an action in virtue of *all* of its morally relevant features. So, the total set of an action’s morally relevant features is relevant to the determination of its being an absolute duty, but only *part* of that set is relevant to the determination of its being a prima facie duty. In this sense, absolute duty is a toti-resultant property and prima facie duty is a parti-resultant one.

There is a *prima facie* duty for a person to perform an action A in virtue of A's possessing feature F =_{df.} there is a moral reason for the person to perform A in virtue of A's possessing F.

(A similar modification to ~PFD is also needed, though I trust that the reader can imagine what such a modification would look like.) From now on, PFD (and ~PFD) should be understood as being modified in this way.

With that, we are now in a position to see what Ross's theory of rightness is in *R&G*.

1.4. Ross's Theory of Rightness

In *R&G*, Ross's theory of rightness is revealed in two separate passages. Here's the first one:

It is worth while to try to state more definitely the nature of the acts that are right. We may try to state first what (if anything) is the universal nature of *all* acts that are right....[R]ight acts can be distinguished from wrong acts only as being those which, of all those possible for the agent in the circumstances, have the greatest balance of *prima facie* rightness, in those respects in which they are *prima facie* right, over their *prima facie* wrongness, in those respects in which they are *prima facie* wrong. (1930, 41)

In this passage, Ross claims to be identifying the "universal nature" of all right actions. The universal nature of right actions, he says, is that they have a greater overall balance of *prima facie* rightness over *prima facie* wrongness than do their alternatives (i.e., the other actions the agent can perform instead). So, then, Ross appears to be endorsing at least the following claim in the passage:

R1: An action A is morally right for S(A)⁹ to perform if and only if S(A) can perform A, and A's total prima facie rightness minus its total prima facie wrongness is greater than that of any other action S(A) can perform instead,

where

An action's *total* prima facie rightness (wrongness) = the sum of the strengths of all of the prima facie duties that there are for the agent to (not to) perform it.

R1 is merely a biconditional: it merely tells us the necessary and sufficient conditions for the moral rightness of actions. However, I think that Ross also accepts a somewhat stronger claim. I think he also accepts the following:

R2: An action A is morally right for S(A) to perform if and only if (and because) S(A) can perform A, and A's total prima facie rightness minus its total prima facie wrongness is greater than that of any other action S(A) can perform instead.

The difference between R1 and R2 is the addition, in R2, of the words "and because." R2 therefore tells us not only what property rightness is coinstantiated with, but it also tells us what *makes* right actions right—it specifies the property *in virtue of which* right actions are right.

That Ross accepts R2, and not just R1, is suggested by the passage just cited. When Ross claims to be stating the "universal nature" of all acts that are right, it's reasonable to interpret this as him telling us what makes all right actions right. It's reasonable, in other words, to interpret him as accepting R2 in the passage. However, the second passage where Ross states his theory of rightness provides an even more conclusive reason for thinking that he accepts R2. Here's the passage:

⁹ "S(A)" refers to the person performing action A (i.e., A's "agent").

We have reached the result that my act is right *qua* being an ensuring of one of the particular states of affairs of which it is an ensuring, viz., in the case we have taken, of my friend's receiving the book I have promised to return to him. But this answer requires some correction; for it refers only to the *prima facie* rightness of my act. If to be a fulfillment of promise were a sufficient ground of the rightness of an act, all fulfillments of promises would be right, whereas it seems clear that there are cases in which some other *prima facie* duty overrides the *prima facie* duty of fulfilling a promise. The more correct answer would be that the ground of the actual rightness of the act is that, of all acts possible for the agent in the circumstances, it is that whose *prima facie* rightness in the respects in which it is *prima facie* right most outweighs its *prima facie* wrongness in any respects in which it is *prima facie* wrong. But since its *prima facie* rightness is mainly due to its being a fulfillment of promise, we may call its being so the salient element in the ground of its rightness. (1930, 46)

Ross says here that the "ground" of the rightness of the action in question is not that it is the fulfillment of any single *prima facie* duty, but rather, that its total *prima facie* rightness over wrongness is greater than that of any of its alternatives. While Ross is explicitly talking about only one action here, it is plausible to assume that he would say similar things about every other right action. After all, there doesn't seem to be anything special about the action Ross is discussing, and he even suggests that any other action with the same ground will likewise be right. So, I think there is good reason to believe that Ross is here identifying what he thinks is the ground of all right actions. He's suggesting, in other words, that the ground of all right actions is that their total *prima facie* rightness over wrongness is greater than that of any of their alternatives. But when Ross talks about the "ground" of a right action, he clearly means to be talking about the reason or explanation for why it is right—that is, the property in virtue of which it is right (see 1930, 10; 1939, 27-28). This passage therefore seems to show that Ross accepts R2 in *R&G*.

If I am correct in thinking that Ross accepts R2, then his theory of rightness is similar in form to Moore's, at least in Moore's later work. In his *Ethics* ([1912] 2005, esp. chap. 2), Moore aims to identify not only the necessary and sufficient conditions of the rightness of actions, but also the ground of rightness. As we've seen, Moore's view is that an action is right if and only if (and because) it would bring about a greater balance of intrinsic goodness over intrinsic badness than any of its alternatives would bring about. Ross, of course, does not accept this, but he does think that the necessary and sufficient conditions of rightness can be identified, as can its ground. And Ross's theory of rightness seeks to identify what, in his opinion, these things are. Ross and Moore are therefore engaged in the same overarching project, though they complete that project in very different ways. This parallel between Ross and Moore should be kept in mind as we move forward, for as we'll see in the next chapter, the failure to recognize it has led many commentators of Ross astray.

A few additional comments on R2 are necessary at this juncture.

First, R2 is Ross's theory of *rightness*. As noted above,¹⁰ Ross uses the term "right" to mean "obligatory" (see 1930, 3-4). I shall therefore refer to R2 as "Ross's Theory of Obligation in *The Right and the Good*," or "RTO_{RG}" for short. From now on, I will generally abstain from using the term "right" since it seems to me to be ambiguous between "obligatory" and "permissible," and when I do use it (see, esp., chapters 7-8), I hope it is clear what I mean.

¹⁰ See note 3.

Second, if R2 is Ross's theory of obligation in *R&G*, then the following are surely his theories of permissibility and wrongness in *R&G*:

RTP_{RG}: An action A is morally permissible for S(A) to perform if and only if (and because) S(A) can perform A, and A's total prima facie rightness minus its total prima facie wrongness is **at least as great as** that of any other action S(A) can perform instead.¹¹

RTW_{RG}: An action A is morally wrong for S(A) to perform if and only if (and because) S(A) can perform A, and A's total prima facie rightness minus its total prima facie wrongness is **less than** that of some other action S(A) can perform instead.

Third, RTO_{RG}-RTW_{RG} make use of the term "action" instead of the term "act."

According to Ross, there is an important difference between acts and actions. He says that an act is "the thing done, the initiation of a change," whereas an action is "the doing of it, the initiation of change, from a certain motive" (1930, 7). And Ross insists that acts, not actions, are the sorts of things that are morally obligatory (permissible, wrong) (7). He would therefore urge us to replace all instances of "action" in RTO_{RG}-RTW_{RG} with "act." However, after making the distinction between acts and actions, Ross often ignores it (see, e.g., 1930, 16). I will follow Ross's lead. I will use the terms "act" and "action" interchangeably; I will use both to mean, more or less, what Ross means by "act."

¹¹ That Ross would accept this is suggested by his (1930, 3-4). Note that RTO_{RG} does not allow for more than one alternative to be obligatory in a situation since only one alternative can have a total balance of prima facie rightness over wrongness that is *greater* than that of every other alternative. However, RTP_{RG} does allow for more than one alternative to be permissible in a situation, for there can be several alternatives that have a total balance of prima facie rightness over wrongness that is *at least as great* as that of any alternative. This will happen when several alternatives are tied for having the greatest total balance of prima facie rightness over wrongness.

It might be helpful at this point for me to say a bit more about Ross's theory of action. In *R&G*, Ross does not develop his account of action in much detail. He merely says, as we've seen, that to act is to "initiate a change." In *FE*, Ross again suggests that to act is "to produce some change in something" (1939, 153), but this time he's a bit more informative. He adds that an action is "the causing of a certain change by setting oneself to cause it" (160). The idea here seems to be this. Suppose I set myself to clap my hands, and this self-exertion causes my hands to clap. In this case, I have performed an action: I have clapped my hands. This action consists of a self-exertion of mine—my setting myself to clap my hands—causing a certain event (i.e., change), namely the clapping of my hands.

I find this rough account of action plausible, though many details remain to be given. Here are few of them:

1. Ross speaks of "self-exertions" and "setting oneself to do something," but what exactly does he mean? He is not very clear about the matter. I suspect that for Ross, a self-exertion to do X or a setting of oneself to do X is simply a decision or an attempt to do X. But more exploration of this suspicion is needed.
2. On Ross's account, an action is an event caused by a self-exertion (decision, attempt). But this needs clarification. Is an action an event that's caused *solely* by a self-exertion? What if the event is caused partly by a self-exertion and partly by other things? Is that an action too? Presumably Ross would say that it is. But what if the causal work done by the self-exertion is minuscule? Or what if the causal work done by the self-exertion merely overdetermines the event? Do we

still have an action then? These are difficult questions that I shall not attempt to answer here.

3. On Ross's theory of action, how are actions to be individuated? This is not a question that Ross discusses, but here's a natural way of answering it. For Ross, an action consists of a self-exertion causing an event. So, we can say that one action, A, is identical to another action, B, iff A and B consist of the same self-exertion and the same event. This, of course, raises questions about how decisions and events are to be individuated. These are also difficult issues that I won't be able to address here.

4. How would Ross distinguish between intentional and unintentional actions?

Although Ross does little to answer this question (though see 1939, 125, 128), here's a plausible proposal: an action is intentional just in case the self-exertion component of the action is a self-exertion to bring about the event component of it. Here's another way to put this. On Ross's view, an action is a complex entity that involves a self-exertion (call it "S") and a causally related event (call it "E"). On the proposed suggestion, an action is intentional iff S is a self-exertion to bring about E. Consider, for instance, the example just given of my clapping my hands. In this example, I set myself to clap my hands, and this self-exertion causes my hands to clap. This action—my clapping my hands—is thus an intentional action: its self-exertion component is a self-exertion to bring about its event component. Suppose, on the other hand, that I clap my hands and this startles someone sitting on the other side of the room. Suppose I don't know that anyone is in the room with me, and so while my self-exertion to clap my hands

causes someone in the room to become startled, it is not a self-exertion to startle anyone. In this case, my action—my startling someone—is unintentional. Of course, the account of intentional action that I am floating here is merely a sketch. Much more will need to be said to make it fully adequate.

5. Ross's theory of action is a causal theory of action. However, causal theories of action are notorious for having difficulty with cases involving causal deviance.¹² So, we may need to modify Ross's theory so that according to it, an action is an event that is caused *in the right way* by a self-exertion. However, whether Ross's theory really needs such modification is unclear to me. Of course, if it does need to be modified in this way, then we'll have to say more about what is meant by being caused "in the right way."

This, of course, is only a selection of the issues that a proponent of Ross's theory of action would ultimately need to address, and a very sketchy attempt at addressing some of them. Still, I think that Ross at least provides us with the foundation of a plausible theory of action, one that it is fruitful to build upon.

Fourth, RTO_{RG} - RTW_{RG} make use of the term "total prima facie rightness (wrongness)." As mentioned above, the total prima facie rightness (wrongness) of an action is the sum of the strengths of all of the prima facie duties that there are for the agent to (not to) perform it. However, this stands in need of clarification, for Ross makes a distinction between fundamental and derivative prima facie duties (1930, 27-28). While he never explicitly says what this distinction amounts to, the idea seems to be this.

¹² See, e.g., Davidson (1973).

Suppose there is a *prima facie* duty for me to keep a promise of mine, but this duty exists only because there is a *prima facie* duty for me to maximize the good, and keeping my promise in this instance maximizes the good. In this case, the *prima facie* duty for me to keep my promise is derivative, not fundamental; it exists only because there is a *prima facie* duty for me to do something else. Suppose, on the other hand, that there is a *prima facie* duty for me to maximize the good, and it's not the case that this duty exists only because there is a *prima facie* duty for me to do something else; rather, the duty exists "on its own" or "in its own right." In this case, the *prima facie* duty for me to maximize the good is fundamental. When Ross talks about *prima facie* duties, he is almost always talking about fundamental *prima facie* duties: he rarely discusses derivative *prima facie* duties.¹³ For this reason, he surely intends to be giving an account of obligation, permissibility, and wrongness in terms of fundamental *prima facie* duties, not derivative ones. So, we should assume that the total *prima facie* rightness (wrongness) of an action is a function of the strengths of the *fundamental* *prima facie* duties that there are for the agent to (not to) perform it. From now on, when I use the term "*prima facie* duty," I will always, unless otherwise noted, be referring to a *fundamental* *prima facie* duty.

Fifth, while there is good evidence that Ross accepts RTO_{RG} in *R&G*, I would be remiss not to point out that he does sometimes suggest a similar, though not equivalent, theory of obligation in the work. Consider, for instance, what he says here:

When I am in a situation, as perhaps I always am, in which more than one of these *prima facie* duties is incumbent on me, what I have to do is to study the situation as fully as I can until I form the considered opinion (it is never more) that in the circumstances one of them is more incumbent than any other; then I am bound to

¹³ Though see 1930, 27-28.

think that to do this *prima facie* duty is my duty *sans phrase* in the situation. (1930, 19; see also 25)

In this passage, Ross may be suggesting something like the following:

R3: An action A is morally obligatory for S(A) to perform if and only if (and because) S(A) can perform A, and there is a *prima facie* duty for S(A) to perform A that is stronger than any of the *prima facie* duties that there are for S(A) to perform any of its alternatives.

RTO_{RG} and R3 do not deliver the same moral verdicts in every case. The following example should make this clear. Suppose I can either perform action A or perform action B. There is a *prima facie* duty of stringency 5 for me to perform A. On the other hand, there are two *prima facie* duties for me to perform B (one of stringency 4, the other of stringency 3). There are no other *prima facie* duties involved in the situation. According to R3, I am obligated to perform A (since there is a *prima facie* duty for me to do it that is stronger than any *prima facie* duty for me to do B). According to RTO_{RG}, on the other hand, I am obligated to perform B (since B's *total* *prima facie* rightness over wrongness is greater than A's). So, the moral verdicts issued by RTO_{RG} and R3 sometimes diverge. Moreover, as this example illustrates, RTO_{RG} is more plausible than R3; surely I am obligated to perform B, not A, in my circumstances. Because of this, I will charitably assume that RTO_{RG} is Ross's theory of obligation in *R&G*, not R3. I think that Ross is simply being a little sloppy in the passage cited above.

We are now almost in a position to see how Ross's theory of rightness solves the problems of Moore's theory and Kant's theory. However, there is one more preliminary task that needs completing: I need to discuss the *prima facie* duties that Ross thinks we have.

1.5. Our Prima Facie Duties, According to Ross

As mentioned above, in *R&G* Ross claims that there are prima facie duties of (1) fidelity, (2) reparation, (3) gratitude, (4) justice, (5) beneficence, (6) self-improvement, and (7) non-maleficence (1930, 21). Ross is careful not to commit himself to this list's being "final" or "complete" (1930, 20). He thus seems willing to admit that further consideration might lead to additions to or subtractions from his list. But he does claim that as far as he can tell, his list captures "all the ways in which *prima facie* duties arise" (27).¹⁴ Let's look at each of these duties in a bit more detail.¹⁵

1.5.1. Fidelity

Ross says that fidelity is the keeping of either an implicit or an explicit promise (1930, 21). He is careful to point out that fidelity is not the *motive* or *intention* to keep a promise; rather, it's the *actual* keeping of a promise, regardless of motive. Ross suggests that in ordinary usage, the term "promise keeping" can refer merely to the motive (or disposition) to keep one's promise. This is why he is, at least at first, careful to say that there is a prima facie duty of *fidelity* instead of saying that there is a prima facie duty of promise keeping (22). However, after making the distinction between fidelity and

¹⁴ As mentioned above, when I talk about prima facie duties, I'm talking about *fundamental* prima facie duties, not derivative ones. I am thus ignoring here duties like the duty to obey the law and the duty to tell the truth, for although Ross thinks that these duties exist, he thinks they are derivative ones (1930, 27-28, 54-55).

¹⁵ In what follows, I will be discussing what Ross says about these duties in *R&G*. In the footnotes, I will say a few things about what Ross says about them in *FE* (see also chapter 4 note 6).

promise keeping, Ross often ignores it and simply uses the term “promise keeping” to mean “fidelity.”¹⁶ I will do the same.

1.5.2. Gratitude

Ross describes gratitude as the return of services received (1930, 3).¹⁷ Ross is again careful to point out that an act of gratitude is the actual return of services received, not merely the motive to do so. Sometimes, Ross claims, the term “gratitude” is used to refer to the latter, not the former. But Ross makes it clear that this is not how he intends to use the term (22-23).

1.5.3. Reparation

Ross says very little about reparation. He simply says that reparation rests “on a previous wrongful act” (1930, 21) and that it arises from “the infliction of injuries on others” (27).¹⁸ Given what he has said about fidelity and gratitude, Ross would surely add that there is a prima facie duty to in fact make reparation for past wrongs, not a prima facie duty to have a reparative motive. But he doesn’t specifically mention this. Perhaps

¹⁶ Indeed, in *FE*, Ross abandons the term “fidelity” altogether and speaks only of promise keeping. Interestingly, in *FE*, Ross seems to suggest that the duty of promise keeping is not a fundamental one but rather is derived from a broader duty not to “fail” or “let down” the promisee (1939, 110). For this reason, Ross holds that there is no prima facie duty to keep a promise that the promisee no longer wants one to keep (110). Ross also holds that there is no prima facie duty to keep a promise that is impossible for one to keep (108-9).

¹⁷ In *FE*, Ross further describes gratitude as a “return [of] good for good” (1939, 102), and a “return of benefits we have received” (289; see also 76, 98).

¹⁸ In *FE*, Ross talks of “making reparation for an injury” (1939, 146; see also 97) and making “compensation to any one for any wrong we have done to him” (76; see also 319).

this is because he doesn't feel the term "reparation" is ordinarily used to refer to any sort of motive.

1.5.4. Beneficence, Self-Improvement, and Justice

Ross says that beneficence is the increase of another person's intelligence, virtue, or pleasure (1930, 21).¹⁹ While he isn't very clear about the matter, we can plausibly take him as holding that intelligence, virtue, and pleasure are the sole constituents of a person's welfare (see also 1930, chap. 5). That is to say, these things (and only these things) are good in themselves for a person. So, I think that we can say, more generally, that for Ross, beneficence is the increase of another person's welfare. In any case, beneficence has nothing to do with motives. To make this clear, Ross says that he intentionally uses the term "beneficence" instead of "benevolence," which he thinks *is* commonly associated with a motive (23, 53).

Ross initially says that self-improvement is the increase of one's own virtue or intelligence (1930, 21). Later, he adds that the increase of one's own pleasure also counts as self-improvement. At first, Ross hesitates to say that the increase of one's own pleasure counts as self-improvement because many would agree that there is a *prima facie* duty of self-improvement but deny that there is a *prima facie* duty to increase one's own pleasure. As he says, "it seems to be a very stubborn fact, that in our ordinary consciousness we are not aware of a duty to get pleasure for ourselves" (1930, 25-26).

¹⁹ Ross would likely add that the increase of a non-human animal's pleasure (virtue?, intelligence?) also counts as an act of beneficence (1930, 49, 138). So, strictly speaking, I should say that for Ross, an act of beneficence is an act that increases the virtue, intelligence, or pleasure of another *thing* (instead of another "person"). However, I will ignore this complication and use the term "person" when I really should use the term "thing."

However, Ross ultimately disagrees with ordinary consciousness on this point;²⁰ he claims that there is in fact a *prima facie* duty to increase one's own pleasure. Why, then, are so many inclined to deny the existence of such a duty? Ross gives two reasons. First, one might notice that we typically don't have a strong impulse to do what we have a *prima facie* duty to do, but we *do* have a strong impulse to increase our own pleasure. In light of this, one might infer that there is no *prima facie* duty to promote our own pleasure (24). Second, one might notice that we generally have to sacrifice our own pleasure to do what there is *prima facie* duty for us to do, but we don't have to sacrifice our own pleasure to increase our own pleasure. In light of this, one might also infer that there is no *prima facie* duty to increase our own pleasure (25). However, Ross seems to think that each of these inferences is fallacious, and he concludes that there is indeed a *prima facie* duty to promote one's own pleasure (24-26).²¹ We should therefore admit, Ross says, that self-improvement involves the increase of not only one's virtue or intelligence,

²⁰ This marks one of the few times that Ross is willing to go against ordinary moral consciousness, and it stands in tension with his general methodology, as it is indicated in the following memorable passages (among others): "I would maintain, in fact, that what we are apt to describe as 'what we think' about moral questions contains a considerable amount that we do not think but know, and that that this forms the standard by reference to which the truth of any moral theory has to be tested, instead of having itself to be tested by reference to any theory" (1930, 40); "The existing body of moral convictions of the best people is the cumulative product of the moral reflection of many generations, which has developed an extremely delicate power of appreciation of moral distinctions; and this the theorist cannot afford to treat with anything other than the greatest respect. The verdicts of the moral consciousness of the best people are the foundation on which he must build; though he must first compare them with one another and eliminate any contradictions they may contain" (41).

²¹ In *FE*, Ross changes his position on this issue; there he claims that there is no *prima facie* duty to increase one's own pleasure. What's more, he says that this view "seems to be so clear as not to need argument" (1939, 273; see also 72). This is a stunning claim given that he held, and argued for, the opposite view in *R&G*. However, I think Ross is wise to change his view on the matter. I'll return to this issue in chapter 6.

but also one's pleasure. But as I noted above, we can plausibly understand Ross as holding that a person's welfare is constituted by virtue, intelligence, and pleasure. We can therefore say that for Ross, self-improvement is simply the increase of one's own welfare.

Ross says that many things are meant by "justice." He says that sometimes people use the term to mean "the payment of debts"; other times, they use it to mean "the reparation of past wrongs" (1930, 26). But this is not how Ross uses the term. By "justice," Ross means "the bringing about of a distribution of happiness between other people in proportion to merit" (26-27, 28).²² When Ross uses the term "happiness" in this context, he is likely referring more generally to a person's welfare. If that's the case, then for Ross, justice amounts to the distribution of welfare in accordance to merit.

While Ross initially says that beneficence, self-improvement, and justice are fundamental *prima facie* duties, he later claims that they "come under the same principle" (1930, 26). By this, he clearly means that these duties can each be derived from a single, more fundamental one. He goes on to describe this duty as the duty to "produce as much good as possible" (27, 35, 39, 154). Ross presumably has in mind here the *prima facie* duty to produce as much *intrinsic* goodness as possible. It is this duty, Ross suggests, that is fundamental, not the duties of beneficence, self-improvement, or justice.

Although Ross suggests that there is a (fundamental) *prima facie* duty to produce as much good as possible, he also often speaks of another, similar *prima facie* duty. He

²² Likewise, in *FE*, he describes the principle of justice as the "moral principle bidding us [to] divide happiness equally between people of equal moral worth." He adds that this same principle "bids us [to] divide it [happiness], so far as we can, unequally between people of unequal moral worth" (1939, 72).

says, “It seems self-evident that if there are things that are intrinsically good, it is *prima facie* a duty to bring them into existence rather than not to do so” (1930, 24). In other places, he speaks of the *prima facie* duty to “bring about the good” (31, 58) and the duty “to promote the general good” (39, 47, 55).²³ In each of these passages, Ross appears to be referring to the same general *prima facie* duty, namely the *prima facie* duty to bring about intrinsic goodness. This duty is obviously very similar to, but not entirely the same as, the *prima facie* duty to produce as much intrinsic goodness as possible. How shall we understand it? Is it yet another fundamental *prima facie* duty? Or is it derived from one of the duties already on Ross’s list?

Ross never explicitly addresses these questions. However, it seems superfluous to hold that there is both a fundamental *prima facie* duty to bring about intrinsic goodness and a fundamental *prima facie* duty to bring about as much intrinsic goodness as possible. Suppose there are two actions available to me in a situation, A and B. A would bring about more intrinsic goodness than B would. If there is a fundamental *prima facie* duty to bring about intrinsic goodness, then there is a fundamental *prima facie* duty to do A in virtue of the fact that it would bring about intrinsic goodness, and there is a fundamental *prima facie* duty to do B in virtue of the fact that it would bring about intrinsic goodness. Moreover, it’s plausible to assume that the strength of the duty to bring about intrinsic goodness is a function of the amount of goodness brought about. So, the *prima facie* duty to do A in virtue of the fact that it would bring about intrinsic goodness is presumably

²³ Similar questions arise in *FE*. In some places in the work, he speaks of a *prima facie* duty to *maximize* the good (1939, 99, 113). In other places, he speaks of a *prima facie* duty to bring about the good (100, 132, 252). He never makes it clear what relation these duties have to each other, if any.

stronger than the prima facie duty to do B in virtue of the fact that it would bring about intrinsic goodness. However, if there is also a fundamental prima facie duty to maximize intrinsic goodness, then there is an *additional* fundamental prima facie duty to do A in virtue of the fact that it maximizes intrinsic goodness. But that would seem to double count the normative significance of A's greater goodness. The normative significance of the fact that A would bring about more goodness than B would was already accounted for by the fact that the duty to do A in virtue of the fact that it would bring about intrinsic goodness is stronger than the duty to do B in virtue of the fact that it would bring about intrinsic goodness. There is thus no need to posit the existence of an additional fundamental prima facie duty to maximize intrinsic goodness.

So, I don't think we should say that there is both a fundamental prima facie duty to bring about intrinsic goodness and a fundamental prima facie duty to maximize intrinsic goodness. But which one should a Rossian accept? The former. In the case just mentioned, a Rossian will surely want to say that there is a fundamental prima facie duty to do B. It would bring about some intrinsic goodness, after all, even if it would bring about less intrinsic goodness than A would. But if there is a fundamental prima facie duty to maximize intrinsic goodness and no fundamental prima facie duty to bring about intrinsic goodness, then since B *doesn't* maximize goodness, it might very well be the case that there is no fundamental prima facie duty to do B. So, I think we should say that there is a fundamental prima facie duty to bring about intrinsic goodness, but no fundamental prima facie duty to maximize intrinsic goodness. This is the view that I will attribute to Ross. Of course, this doesn't preclude Ross from saying that there is a prima facie duty to maximize intrinsic goodness; it just precludes him from saying that the duty

is *fundamental*. He can happily maintain that there is a duty to maximize goodness, but it is a derivative one.

1.5.5. Non-Maleficence

Ross claims that the prima facie duty of non-maleficence can be “summed up under the title of ‘not injuring others’” (1930, 21), and he later describes it as the duty “not to harm others” (22).²⁴ Ross also points out that the prima facie duty of non-maleficence is different from the other duties discussed so far. All of the prima facie duties mentioned thus far have been duties *to* do something (*to* keep promises, *to* return for services received, *to* repair for past wrongs, etc.). However, the duty of non-maleficence is not a duty *to* do anything at all. Instead, it is a duty *not* to do something (namely, *not* to harm another) (21-22).

Notice that Ross only says that there is a prima facie duty not to harm *others*. He doesn’t say that there is a prima facie duty not to harm *ourselves*. This is odd since, we’ve seen, he thinks that there is not only a prima facie duty to improve others (beneficence), but also a duty to improve ourselves (self-improvement). But if he thinks that, then why wouldn’t he likewise hold that just as there is a prima facie duty not to harm others, there is also a duty not to harm ourselves? Perhaps this is merely an oversight on Ross’s part.

²⁴ Ross rarely discusses non-maleficence in *FE*. Indeed, the term never even appears in the work. However, he does, at one point, acknowledge that an act likely to harm someone is prima facie wrong (1939, 89).

1.5.6. Promise Breaking, Ingratitude, Unreparation, Injustice?

Notably absent from Ross's list of prima facie duties are the duties against promise breaking, ingratitude, unreparation, and injustice. Like the prima facie duty of non-maleficence, these duties would presumably be duties *not* to do something, namely not to break one's promises, not to commit acts of ingratitude, not to commit acts of unreparation, and not to commit acts of injustice.²⁵ Ross does, in one instance, suggest that there is a prima facie duty not to break one's promises (1930, 28),²⁶ and I see no reason to doubt that he would likewise admit that there is a prima facie duty not to perform acts of ingratitude, unreparation, or injustice. I shall therefore add these duties to Ross's list, though it must be admitted that he never discusses them explicitly.

1.5.7. Bringing about Intrinsic Badness?

Although Ross never discusses it, if he holds, as I've suggested, that there is a fundamental prima facie duty to bring about intrinsic goodness, then he would surely also admit that there is a fundamental prima facie duty not to bring about intrinsic badness. And if he believes that the prima facie duties of beneficence, justice, and self-improvement are reducible to the prima facie duty to bring about intrinsic goodness, then he would surely also admit that the prima facie duties of non-maleficence, injustice, (and

²⁵ Perhaps Ross would say that an act of ingratitude is the return of bad for good, an act of unreparation is the exacerbation of past wrongs, and an act of injustice is the distribution of happiness disproportionately to merit (i.e., the giving of happiness to the vicious or unhappiness to the virtuous).

²⁶ In *FE*, Ross also suggests that it is prima facie wrong to break a promise (1939, 85, 93, 173).

self-harm?²⁷) are reducible to the prima facie duty not to bring about intrinsic badness. I shall thus attribute this view to Ross.

1.5.8. Summary

In conclusion, I have claimed that it is plausible to hold that Ross's considered view about our prima facie duties is that there is a prima facie duty to (1) keep our promises, (2) perform acts of gratitude, (3) perform acts of reparation, and (4) bring about intrinsic goodness; on the other hand, there is a prima facie duty not to (1) break our promises, (2) perform acts of ingratitude, (3) perform acts of unreparation, and (4) bring about intrinsic badness.²⁸

1.5.9. Loose Ends

A few loose ends need to be tied at this point.

First, in the foregoing, I have attributed to Ross the claim that there is a prima facie duty of promise keeping (reparation, gratitude, etc.), the claim that there is a prima facie duty to keep our promises (perform acts of reparation, gratitude, etc.), and the claim that promise keeping (reparation, gratitude, etc.) is a prima facie duty. Ross often speaks in these varied ways. However, I suspect that it is all somewhat loose talk. What Ross really means to say, I think, is that necessarily, for any action, if it is a promise keeping (act of reparation, gratitude, etc.), then in virtue of that fact, there is a prima facie duty for

²⁷ As mentioned above, Ross never claims that there is a prima facie duty not to harm ourselves.

²⁸ In *FE*, Ross's list of prima facie duties looks very much like the positive portion of this list (1939, 186, 271).

the agent of the action to perform it.²⁹ Of course, that is a mouthful, and so it is convenient to simply say that promise keeping (gratitude, reparation, etc.) is a *prima facie* duty, that there is a *prima facie* duty to keep one's promises (perform acts of gratitude, reparation), and so forth. I shall continue to use loose phrases like these for ease of exposition.

Second, Ross claims that little can be said about the comparative stringencies of the *prima facie* duties on his list. For instance, he says,

For the estimation of the comparative stringency of these *prima facie* obligations no general rules can, so far as I can see, be laid down. We can only say that a great deal of stringency belongs to the duties of "perfect obligation"—the duties of keeping our promises, of repairing wrongs we have done, and of returning the equivalent of services we have received. For the rest, ["the decision rests with perception"³⁰]. (1930, 41-42)

What Ross seems to mean here is that we can't say that, for instance, the strength of the *prima facie* duty of, say, promise keeping is always stronger than the *prima facie* duty of beneficence. In many circumstances it *will* be stronger. But in other circumstances it might not. Similar things apply to the relationship between the other *prima facie* duties on Ross's list: each might outweigh the others in certain situations—there is no lexical ordering of the *prima facie* duties to be found. As modern day philosophers tend to put it, *prima facie* duties are "overridable."³¹ For this reason, there won't be any simple, easily identifiable principle that we can use to determine the relative stringencies of the *prima*

²⁹ That Ross accepts this is implicit throughout chapter 2 of *R&G*, but see especially pp. 32-34.

³⁰ In the text, the phrase "the decision rests with perception" appears in Greek (Ross is quoting Aristotle). Ross's English translation of the phrase appears in a footnote.

³¹ The overridability of *prima facie* duties is implicit throughout chapter 2 of *R&G*.

facie duties involved in most scenarios. To make such a determination, we have no choice but to use our best judgment.

However, while Ross denies that there is any lexical ordering of the *prima facie* duties, he also sometimes suggests that some classes of *prima facie* duties are more stringent than others. For instance, he says that non-maleficence is more stringent than beneficence:

No doubt to injure others is incidentally to fail to do them good; but it seems to me clear that non-maleficence is apprehended as a duty distinct from that of beneficence, and as a duty of a more stringent character. (1930, 21)

But even when we have come to recognize the duty of beneficence, it appears to me that the duty of non-maleficence is recognized as a distinct one, and as *prima facie* more binding. We should not in general consider it justifiable to kill one person in order to keep another alive, or to steal from one in order to give alms to another. (22)

These passages are puzzling. Ross surely doesn't mean that the duty of non-maleficence is *always* stronger than the duty of beneficence. What, then, could he mean when he says that non-maleficence is "more stringent" or "*prima facie* more binding" than beneficence?

Here's a suggestion. When Ross says that non-maleficence is more stringent (or *prima facie* more binding) than beneficence, he means that when an action would bring about a certain amount of good for others and an equal amount of bad (harm) for others, the strength of the *prima facie* duty the agent has not to perform it (in virtue of its being an act of maleficence) is stronger than the *prima facie* duty the agent has to perform it (in virtue of its being an act of beneficence). So, the strength of the duty of non-maleficence will override the strength of the duty of beneficence unless the action in question benefits

others more (perhaps considerably more) than it harms them. In this sense, the duty of non-maleficence is more stringent (or *prima facie* more binding) than the duty of beneficence.

Ross is also sometimes interpreted as holding that promise keeping is more stringent/*prima facie* more binding than beneficence.³² He says,

We...think...that normally promise-keeping, for example, should come before benevolence, but that when and only when the good to be produced by the benevolent act is very great and the promise comparatively trivial, the act of benevolence becomes our duty. (19)

However, the claim that promise keeping is more stringent than beneficence is considerably more difficult to make sense of than the claim that non-maleficence is more stringent than beneficence. I suggested that Ross's claim that non-maleficence is more binding than beneficence amounts to the claim that when an action would bring about equal amounts of benefit and harm to others, the duty of non-maleficence is stronger than the duty of beneficence. This, of course, assumes that it makes sense to talk of "equal amounts" of benefit and harm. It assumes, that is, that we can rank the size of benefits and the size of harms on a single cardinal scale. And while some might challenge this assumption, it is more plausible than the assumption that we can rank the size of *promise keepings* and the size of benefits to others on a single cardinal scale. As we'll later see, Ross thinks that the strength of the duty of promise keeping is a function of its solemnity, its recency, and its importance to the promisee (see sec. 2.4). Perhaps, then, he'd say that the "size" of a promise is some function of these things. But it's difficult to see how we

³² Both McNaughton (1996) and Robinson (2010) interpret Ross this way.

could coherently rank solemnity, recency, importance to a promisee, or any function of these things on the same cardinal scale as the scale we use to measure the size of beneficence. And so I doubt that we can make sense of the claim that promise keeping is more stringent than beneficence in the same way that I suggested we make sense of the claim that non-maleficence is more stringent than beneficence.

How, then, should this claim be understood? I doubt that there is a good answer to this question.³³ However, it's also not clear that Ross really makes the claim in the first place. After all, in the above passage where he allegedly makes the claim, he merely says that "normally promise-keeping, for example, should come before benevolence." So, he may simply be making a statistical generalization about the strengths of these duties in the actual world. That is, he may be suggesting that in the actual world, when there is a conflict between promise keeping and beneficence, promise keeping typically wins—the duty of promise keeping doesn't *always* outweigh the duty of beneficence, but more than not, it does. And while this claim may be false (I won't speculate about the matter here), there is nothing especially mysterious about it.³⁴

Third, Ross thinks it is self-evident that we have the prima facie duties on his list (1930, 12, 20n, 24, 40). He thinks it's self-evident that, say, there is a prima facie duty of promise keeping in just the same way that it's self-evident that $2+2=4$ (29, 32-33). However, Ross thinks that we never know what our *actual* duty is in any realistic

³³ For some recent, unsuccessful attempts, see McNaughton (1996) and Robinson (2010).

³⁴ Ross is sometimes also interpreted as holding that gratitude and reparation are more stringent than beneficence (see, e.g., McNaughton (1996) and Robinson (2010)). However, the textual grounds for attributing these claims to Ross (namely, 1930, 41-42) are even weaker than the grounds for attributing to him the view that promise keeping is more stringent than beneficence.

circumstance we encounter (30-31, 41-42). This is because he thinks, as we've seen, that we rarely know the relative stringencies of the *prima facie* duties involved in any given scenario. Moreover, he notes, our alternatives will always have many unknown consequences. Thus, Ross says, our actions will "have a *prima facie* rightness or wrongness of which we know nothing" (31). For these reasons, Ross thinks that we are always in the dark about our actual duty in any actual situation.³⁵ (Of course, Ross would still admit that we can determine our actual duty in hypothetical situations where we stipulate that our actions have no unknown consequences, and we stipulate the strengths of the *prima facie* duties involved.)

With that, we're now finally in a position to see why Ross's theory avoids the problems of Moore's Ideal Utilitarianism and Kantianism.

1.6. Ross's Theory Compared with Ideal Utilitarianism and Kantianism

Return to the case that vexed Ideal Utilitarianism. In that case, I can either keep my promise and thereby bring about 1,000 units of goodness, or break my promise and thereby bring about slightly more goodness (1,001 units). (There are no other morally relevant features of the case.) Ideal Utilitarianism implies, counterintuitively, that I should break my promise. Ross's theory, in contrast, seems to imply that I ought to keep my promise. Although Ross admits that there is a *prima facie* duty to bring about intrinsic goodness, he also thinks that there is a *prima facie* duty to keep our promises and a *prima facie* duty not to break our promises. But it's presumably the case that the strength of the *prima facie* duty to keep my promise along with the strength of the *prima*

³⁵ I'll return to this claim in sec. 4.5.

facie duty not to break my promise is more than enough to outweigh the strength of the prima facie duty to bring about one additional unit of goodness. If that's the case, then although keeping my promise will bring about slightly less goodness than breaking my promise will, keeping my promise will nonetheless have a greater total balance of prima facie rightness over wrongness. Ross's theory will therefore imply that I ought to keep my promise.

To see this more clearly, consider the following table, which represents my options in the present case and a plausible way of filling in the strengths of the prima facie duties involved:

Table 1: Keeping a slightly sub-optimal promise

Option 1: Bring about 1,000 units of goodness; keep promise		Option 2: Bring about 1,001 units of goodness; break promise	
<u>PF Duty (To)</u>	<u>Strength</u>	<u>PF Duty (To)</u>	<u>Strength</u>
Promise keeping	100	Promise keeping	0
Gratitude	0	Gratitude	0
Reparation	0	Reparation	0
Intrinsic goodness	1,000	Intrinsic goodness	1,001
<u>PF Duty (Not To)</u>	<u>Strength</u>	<u>PF Duty (Not To)</u>	<u>Strength</u>
Promise breaking	0	Promise breaking	100
Ingratitude	0	Ingratitude	0
Unreparation	0	Unreparation	0
Intrinsic badness	0	Intrinsic badness	0
Total PF To – Total PF Not To	1,100 – 0 = 1,100	Total PF To – Total PF Not To	1,001 – 100 = 901

If the strengths of the prima facie duties of my options are as indicated in this chart, then it will follow from RTO_{RG} - RTW_{RG} that I ought to keep my promise in this case, and it's wrong for me to break my promise.³⁶

³⁶ A few comments about this chart:

First, this chart—and many others in this dissertation—assumes that the strengths of prima facie duties can be assigned precise values. Of course, it will be very difficult, if not impossible, for us to say with confidence what precise value a prima facie duty has in any given situation. But I am assuming that it is in principle possible to represent the strength of a prima facie duty in a situation with a number.

Second, this chart—and many others to follow—not only assumes that the strengths of prima facie duties can be assigned precise values, but it also assumes that these values are commensurable. That is, it assumes that there is a single cardinal scale that can be used to measure the strength of any prima facie duty in any situation. The values listed in the “strength” column of the above chart represent the strengths of the prima facie duties when they are ranked on this single cardinal scale. Without a single cardinal scale to measure the strengths of prima facie duties, it would make little sense to add or subtract the values of different prima facie duties. Without such a scale, the value in the final row of the above chart would have no more

As we've seen, Kant would agree with these verdicts. However, he would go even further and say that since the duty of promise keeping is a perfect duty, it is never permissible to break a promise. But this position is highly implausible. It seems permissible for me to break my promise to, say, pick you up at the airport if by doing so I could save someone's life. Ross's theory, however, seems to make the correct conclusions about such a case. He'd say that while there is a *prima facie* duty of promise keeping and a *prima facie* duty against promise breaking, there is also a *prima facie* duty to bring about intrinsic goodness and a *prima facie* duty not to bring about intrinsic badness. And, he'd continue, the strengths of these latter two duties easily outweigh the strengths of the former two in this case; thus, breaking my promise is what I ought to do.

More concretely, here's a plausible way of filling in the strengths of the *prima facie* duties in the case:

significance than the value we get when we add the length of one thing, measured in inches, with the length of another thing, measured in centimeters.

Third, one might wonder why I chose the particular numbers that I did in this case. For instance, why did I assign 100 to the strength of the duty of promise keeping? Why not 99, or 98, or 97, etc.? The answer is that my decision to assign 100 to the strength of promise keeping in this case is largely arbitrary. All that matters for Ross's purposes is that the combined strength of the duty to keep a promise and the duty not to break a promise is greater than the difference between the strength of the duty to bring about intrinsic goodness and the strength of the duty not to bring about intrinsic badness. I have offered one plausible way of filling in the above chart that is consistent with this constraint, though there are certainly many other plausible ways of filling it in as well. In many of the charts that follow, there will likewise be a certain amount of arbitrariness to the numbers I assign to the strengths of *prima facie* duties. There will typically be many plausible ways of filling in the charts contained in this dissertation. I will leave it to the reader to decide whether my way of filling in these charts is in fact a *plausible* way of filling them in.

Table 2: Breaking a promise to save a life

Option 1: Keep promise; allow someone to die		Option 2: Break promise; save a life	
<u>PF Duty To</u>	<u>Strength</u>	<u>PF Duty To</u>	<u>Strength</u>
Promise keeping	100	Promise keeping	0
Gratitude	0	Gratitude	0
Reparation	0	Reparation	0
Intrinsic goodness	0	Intrinsic goodness	1,000
<u>PF Duty Not To</u>	<u>Strength</u>	<u>PF Duty Not To</u>	<u>Strength</u>
Promise breaking	0	Promise breaking	100
Ingratitude	0	Ingratitude	0
Unreparation	0	Unreparation	0
Intrinsic badness	1,000	Intrinsic badness	0
Total PF To – Total PF Not To	100 – 1,000 = -900	Total PF To – PF Not To	1,000 – 100 = 900

If the strengths of the prima facie duties involved in the case are as indicated here, $RTO_{RG}-RTW_{RG}$ implies that I ought to break my promise and save a life, and it's wrong for me to keep my promise and allow someone to die. Ross's theory therefore makes the intuitively correct conclusions about the case.

For these reasons, Ross's theory appears to be an attractive alternative to Ideal Utilitarianism and Kantianism. However, his theory is not without its problems. In chapters 4-6, I will consider several objections to his theory. Prior to doing that, however, I'd like to correct some common ways of misunderstanding Ross's theory. In the next two chapters, I will turn to that task.

CHAPTER 2

ROSS AND PARTICULARISM

2.1. Introduction

Ross is commonly taken to be a *generalist* about prima facie duty. That is, he is commonly thought of as holding that there are some true moral principles having to do with prima facie duty. However, Ross is typically taken to have a very different view about absolute duty. About it, Ross is often considered to be not a generalist, but a staunch *particularist*. That is, Ross is often thought of as denying that there are any true absolute moral principles.¹ Sean McKeever and Michael Ridge, for instance, say the following about Ross:

While Ross denied that one could spell out principles that would determine when one had a duty overall to do something, he insisted that one could specify principles of prima facie duty, i.e., principles operating at the contributory [i.e., prima facie] level. (2008, 1183-84; see also 2005, 85n)

Philip Stratton-Lake says something similar:

There are, he [Ross] maintains, no universal moral laws of the form, “actions of type F are obligatory.” There are, however, laws of the form “actions of type F *tend* to be obligatory.” *Prima facie* duties are these laws. (2000, 82)

Consider also what Jonathan Dancy says in the following passage:

¹ The terms “particularism” and “generalism” are used in many different ways by many different philosophers. As I will be using the terms in this chapter, they refer to the acceptance (or denial) of true moral principles: generalism is the view that there are some such principles; particularism is the view that there are none. For more on the many ways that particularism and generalism have been understood in the literature, see McKeever and Ridge (2005, 2006).

Ross's view is that there are many codifiable principles. Unlike the Welfarist's Sole Principle [i.e., utilitarianism] they are all principles of prima facie duty, not of duty proper. (2004b, 11)

And finally, Pekka Väyrynen claims that "pluralists" such as Ross "defend only contributory [i.e., prima facie] principles and deny the existence of true overall principles" (2006, 717n33; see also 2009, 110-11n).

I agree with these authors when they suggest that Ross accepts prima facie moral principles (and is thus a generalist about prima facie duty). Ross clearly thinks that there are some true prima facie moral principles. He holds, for instance, that there is a prima facie duty to keep one's promises, a prima facie duty to perform acts of gratitude, a prima facie duty not to harm others, and so forth, for the rest of the prima facie duties on his well-known list (1930, 21). However, I take issue with these authors when they suggest that Ross denies that there are any true absolute moral principles (and is thus a particularist about absolute duty). In this chapter, I will explain why.

2.2. Utilitarianism and Kantianism

While Ross is considered to be an arch-particularist about absolute duty, utilitarians are considered to be arch-*generalists* about this sort of duty. Hedonistic act utilitarians, for instance, believe that an action is morally obligatory if and only if it would bring about a greater balance of pleasure over pain than any of its alternatives would bring about. Hedonistic rule utilitarians accept something similar, though they would say that an action is morally obligatory iff it is required by a rule the general adherence to which would maximize the balance of pleasure over pain. Of course, there are many other types

of utilitarian, but they all seem to accept something along these lines, and they all thus seem to accept at least one absolute moral principle.

Kantians, too, are taken to be generalists about absolute duty. Kantians accept claims such as the following: an action is morally permissible iff its maxim can be willed to be a universal law; an action is morally permissible iff its agent, in performing the action, treats humanity as an end in itself, not as a mere means; and an action is morally permissible iff it conforms to the maxims of a member giving universal laws for a merely possible kingdom of ends. There is, of course, much debate among Kant scholars about how to understand these claims and whether they amount to the same thing, as Kant suggests. But for my purposes, the point is that Kantians clearly seem to accept absolute moral principles, however those principles are ultimately understood.

Similar things can be said about divine command theorists, contractualists, and egoists. They all accept propositions of this form: an action is morally obligatory (permissible, wrong) iff _____. They all therefore seem to accept absolute moral principles, and they are thus considered to be paradigmatic examples of generalists about absolute duty.

However, it seems to me that if we consider utilitarians, Kantians, etc. to be generalists about absolute duty, then we should consider Ross to be a generalist about this sort of duty too. After all, as we've seen, there is good reason to hold that Ross accepts

R1: An action A is morally right for S(A) to perform if and only if S(A) can perform A, and A's total prima facie rightness minus its total prima facie wrongness is greater than that of any other action S(A) can perform instead.

So, like utilitarians and Kantians, Ross seems to accept a proposition of this form: an action is morally obligatory (permissible, wrong) iff _____. If this makes utilitarians and Kantians generalists about absolute duty, then surely it makes Ross a generalist about it as well.

Since utilitarians and Kantians are so often taken to be generalists about absolute duty while Ross is so often taken to be a particularist about it, one may suspect that I have misunderstood what an absolute moral principle is. But as I will show in the remainder of this chapter, other plausible ways of conceiving of absolute moral principles don't do much better at making utilitarians and Kantians generalists about absolute duty while also making Ross a particularist about it. My general conclusion, then, is that Ross is no more a particularist about absolute duty than a utilitarian or a Kantian is. While this conclusion is interesting in its own right, it is also important, I will argue, because it prevents us from overlooking Ross's theory of moral obligation and because it may have implications on the broader debate between particularists and generalists.

2.3. What's an Absolute Moral Principle?—Some Initial Answers

In the previous section, I was implicitly relying on something like the following conception of absolute moral principles:

MP1: An absolute moral principle is, by definition, a proposition that has the following form: an action is morally obligatory (permissible, wrong) iff _____.

As we've seen, if we accept this account of absolute moral principles, then we'll have to admit that not only do utilitarians and Kantians accept absolute moral principles, but Ross accepts them too.

Of course, I may have misunderstood what an absolute moral principle is. Perhaps MP1 is false. But if it's false, then what should we replace it with? Well, one thing that particularists are fond of saying is that moral principles must be explanatory.²

Particularists typically insist that moral principles be explanatory in order to prevent the supervenience of the moral on the natural from generating moral principles. I take it that to say that a moral principle is explanatory is, roughly, to say that the principle specifies the feature or features in virtue of which some moral property obtains. Or, to put it another way, it's to say that the principle specifies what makes it the case that a certain moral property is instantiated.

But notice that according to MP1, a proposition might be a moral principle without being explanatory. For instance, the following propositions count as absolute moral principles on MP1: an action is morally obligatory iff it is morally obligatory; an action is morally obligatory iff it is morally required; and an action is morally obligatory iff it is morally wrong not to perform. But no particularist (or generalist) would regard these propositions as absolute moral principles. (If they did, generalism about absolute duty would obviously be true.) And, I suspect, they'd say that these propositions are not moral principles because they are not explanatory. They don't tell us *why* obligatory actions are obligatory.

² See, for instance, Lance and Little (2006a, 307-9), Shafer-Landau (1997, 584-86), McNaughton (1988, 191-92), and McKeever and Ridge (2005, 85; 2006, 8).

This suggests the following modification of MP1:

MP2: An absolute moral principle is, by definition, a proposition that has the following form: an action is morally obligatory (permissible, wrong) iff **and in virtue of the fact that** _____.

But notice that R1 doesn't count as an absolute moral principle on MP2. After all, R1 is merely a biconditional. It merely says that moral obligation is necessarily coinstantiated with the property of having the greatest balance of total prima facie rightness over total prima facie wrongness. It doesn't make the stronger claim that obligatory actions are obligatory *because* they have this property. Thus, if MP2 is true, Ross's acceptance of R1 does not show that he accepts absolute moral principles.

Furthermore, it seems that utilitarians and Kantians do accept absolute moral principles if MP2 is correct. As I've said, hedonistic act utilitarians hold that an action is morally obligatory iff it would bring about a greater balance of pleasure over pain than any of its alternatives would bring about, and they will surely also hold that obligatory actions are obligatory in virtue of this fact. The same goes for other utilitarians: the biconditionals they accept will surely be considered by their proponents to be explanatory. Likewise, Kantians accept several biconditionals—one having to do with the universalizability of maxims, another having to do with treating humanity as an end in itself, and another having to do with a merely possible kingdom of ends—and they will, I assume, hold that at least one of these biconditionals is explanatory.³

³ Mark Timmons (2002, 163; 2006, 190), for instance, explicitly says that the second of these biconditionals (the so-called "Principle of Humanity") specifies, for Kant, what makes permissible actions permissible.

So, one might think that MP2 is just what we are looking for: it's a conception of moral principles that makes Ross a particularist about absolute duty and utilitarians and Kantians generalists about it.

But the matter is not so simple, for as we've seen, there's good reason to believe that not only does Ross accept R1, but he also accepts

RTO_{RG}: An action A is morally right for S(A) to perform if and only if (and because) S(A) can perform A, and A's total prima facie rightness minus its total prima facie wrongness is greater than that of any other action S(A) can perform instead.

The "and because" in RTO_{RG} is simply another way of saying "and in virtue of the fact that." RTO_{RG} therefore counts as a moral principle on MP2. So, if we accept this account of moral principles, we should conclude that like utilitarians and Kantians, Ross accepts absolute moral principles and should thus be regarded as a generalist, not a particularist, about absolute duty.

However, even if RTO_{RG} is to some degree explanatory, as Ross maintains, it may seem trivial in a way that utilitarian and Kantian principles do not. After all, as discussed in sec. 1.3, Ross's notion of a prima facie duty is often understood in terms of the concept of a moral reason. More precisely, many philosophers assume that to have a prima facie duty to (not to) perform an action just is to have a moral reason to (not to) do it. But if we understand the notion of a prima facie duty in this way, then RTO_{RG} will amount to the idea that an obligatory action is one that has the greatest balance of moral reasons for it minus moral reasons against it. But this, one might think, is hardly controversial: surely we should do what we have most moral reason to do. The same cannot be said,

however, about utilitarians and Kantians: the claims they accept seem to be much more substantive.

Consider, then, the following modification of MP2:

MP3: An absolute moral principle is, by definition, a **non-trivial** proposition that has the following form: an action is morally obligatory (permissible, wrong) iff and in virtue of the fact that _____.

Will this conception of absolute moral principles render Ross a particularist about absolute duty and utilitarians and Kantians generalists about it?

The answer, I think, is “no.” RTO_{RG} is not as trivial as it may first appear. In fact, several philosophers reject it. For instance, Ned Markosian appears to reject it in his (2009). There he argues that a view he calls “Rossian Minimalism” is “the best ethical theory that can be stated in terms of Ross’s notion of a *prima facie* duty” (9). Rossian Minimalism is the view that an action is morally obligatory iff its *prima facie* wrongness is less than that of any of its alternatives.⁴ Rossian Minimalism is thus very different from RTO_{RG} since according to RTO_{RG} , moral obligation is determined by both *prima facie* wrongness *and* *prima facie* rightness—in particular, obligatory actions *maximize* the balance of total *prima facie* rightness over wrongness. But Rossian Minimalism is not a maximizing theory. According to it, obligation is determined only by *prima facie* wrongness—obligatory actions are ones that are *least* *prima facie* wrong. Markosian

⁴ Markosian (2009, 7) officially states Rossian Minimalism in terms of permissibility, not obligation. However, I assume he’d have no problem with my statement of the view.

argues that instead of accepting a maximizing theory such as RTO_{RG} , we should accept a minimizing one like Rossian Minimalism.⁵

Other philosophers also reject RTO_{RG} , at least if we understand *prima facie* duties in terms of moral reasons. Doug Portmore (2008; 2011, chap. 5) is one such philosopher. Portmore is interested in finding a moral theory that accords as closely as possible with commonsense morality. And he thinks that supererogatory actions and agent-centered options⁶ are part of commonsense morality. Portmore argues that in order to accommodate these things, we must admit that the deontic⁷ status of an action is determined not only by moral reasons, but also by *non*-moral reasons. If Portmore is correct about this, then RTO_{RG} is false if it amounts to the idea that an action is obligatory just in case the balance of *moral* reasons for it minus the *moral* reasons against it is greater than that of any of its alternatives. RTO_{RG} , Portmore would urge, needs to be revised so that obligation is a function of not only the agent's moral reasons, but her non-moral reasons as well.

Jonathan Dancy (2004a; 2004b, chap. 2) and Joshua Gert (2003; 2004, chap. 2) suggest something similar. They draw a distinction between requiring reasons and

⁵ I will discuss Rossian Minimalism in much greater detail in chapter 5.

⁶ According to Portmore (2011, 95, 237), an agent-centered option is an option either to do what makes things better overall but worse for oneself, or to do what makes things better for oneself but worse overall.

⁷ As I'll use the term, for something to have "deontic status" is for it to be either obligatory, permissible, or wrong.

justifying reasons.⁸ Requiring reasons are, roughly speaking, ones that can make an otherwise merely permissible action obligatory. Justifying reasons, on the other hand, are ones that cannot make an otherwise merely permissible action obligatory, but they *can* make an otherwise wrong action permissible. If we understand prima facie duties in terms of moral reasons, we should clearly understand them in terms of requiring reasons. This is because according to Ross (1930, 19), if there is a prima facie duty to perform an action, this means that the action would be obligatory if it had no other morally relevant properties. But only requiring reasons have this feature. So, if we understand RTO_{RG} in terms of moral reasons, it will amount to the claim that an action is morally obligatory just in case the balance of moral requiring reasons in favor of it minus the moral requiring reasons against it is greater than that of any of its alternatives. But as Dancy and Gert suggest, this is problematic: moral obligation seems to be a function of not only requiring reasons, but a function of justifying reasons as well.⁹

Of course, there is more to be said about all of this—perhaps the philosophers I’ve mentioned are too quick to reject RTO_{RG} . However, this is not an issue that I need to discuss any further here. My point is merely that RTO_{RG} is not as obvious or trivial as it may first appear; in fact, several philosophers give strong considerations against it. MP3 is therefore not the conception of absolute moral principles that we are looking for: both

⁸ This is Gert’s terminology. Dancy uses the terms “peremptory reason” and “enticing reason” instead of “requiring reason” and “justifying reason,” respectively. Dancy, however, seems to have the same distinction in mind as the one Gert does.

⁹ I will return to this idea, as well as the idea suggested by Portmore in the previous paragraph, in chapter 6.

Ross and utilitarians/Kantians will be generalists about absolute duty according to it. The search continues.

2.4. What's an Absolute Moral Principle?—Another Answer

One might think that the following is an important difference between utilitarians and Ross. Utilitarians think that obligatory actions are obligatory in virtue of possessing a *descriptive* feature. In the case of hedonistic act utilitarianism, this feature is the feature of bringing about a greater balance of pleasure over pain than any alternative would bring about. In the case of hedonistic rule utilitarianism, this feature is the feature of being required by a rule the general adherence to which would maximize the balance of pleasure over pain. Ross, on the other hand, thinks that obligatory actions are obligatory in virtue of possessing a *normative* feature: namely, in virtue of having a greater balance of prima facie rightness over wrongness than that of any alternative. Perhaps this is what makes utilitarians generalists about absolute duty but Ross a particularist about it.

More precisely, suppose we modify MP3 as follows:

MP4: An absolute moral principle is, by definition, a non-trivial proposition that has the following form: an action is morally obligatory (permissible, wrong) iff and in virtue of the fact that _____, **where this blank contains only descriptive terms.**¹⁰

¹⁰ Lots of particularists and generalists suggest that moral principles link the descriptive and the normative (see, for instance, Little 2000; McNaughton 1988, 190-92; Leibowitz 2009, 184-85; Cullity 2002, 170-71; and Jackson, Pettit, and Smith 2000, 80-81). There's an interesting question about how to characterize the difference between descriptive and normative terms. I shall not pursue this question here. I will assume that the distinction is clear enough.

If we adopt this, then Ross's acceptance of RTO_{RG} does not show that he accepts absolute moral principles. However, it might be thought that utilitarians and Kantians do accept absolute moral principles as defined by MP4; so, if we accept it, we'll be able to say that utilitarians and Kantians are generalists about absolute duty without having to say that Ross is a generalist about it as well.

Again, however, the issue is more complicated than it might initially appear. First of all, it's not clear that all utilitarians will be committed to generalism about absolute duty if we accept MP4. Consider G. E. Moore, for instance. As we've seen, according to his version of utilitarianism (Ideal Utilitarianism), an action is right iff (and because) it would bring about a greater balance of intrinsic goodness over intrinsic badness than any of its alternatives would bring about. However, this won't count as an absolute moral principle on MP4 since its right-hand side contains normative terms ("goodness" and "badness"). So accepting this won't, by itself, make Moore a generalist about absolute duty. But that seems odd: Moore is supposed to be a paradigmatic example of a generalist about this kind of duty, and surely it's his acceptance of Ideal Utilitarianism that makes him one.

It's also not clear to me that Kantians will necessarily be committed to generalism about absolute duty if we accept MP4. Whether they will be depends on how we understand the key notions that Kantians appeal to. Consider, for instance, what it means to say that someone "treats humanity as an end in itself." Some have thought that treating humanity as an end in itself involves treating people with dignity and respect. Others have thought that it involves treating people with equal moral consideration. Still

others have thought that it involves recognizing the value of humanity in people.¹¹ Of course, the concepts of dignity, respect, moral consideration, and value are normative concepts. But if the notion of treating humanity as an end in itself is, in part, a normative notion, then by accepting that an action is morally permissible iff it treats humanity as an end in itself, a Kantian won't necessarily be committed to generalism about absolute duty, at least if MP4 is true. Similar things can be said if either the notion of universalizability or the notion of a universal law in the kingdom of ends turns out to be normative.

This is not the place to discuss the proper way to understand these Kantian concepts. My point is merely that it's not obvious that Kantians will be committed to generalism about absolute duty if MP4 is true. Whether they are depends on some tricky interpretive matters. But one might again find this odd: Kantians are also supposed to be paradigmatic examples of generalists about absolute duty, and they should be so considered irrespective of the finer details of Kant exegesis.

Moreover, I see no reason to doubt that Ross would have accepted absolute moral principles even if MP4 is true. To see this, let's take a closer look at Ross's moral theory. As we've seen, he thinks that obligatory actions are ones that have the greatest balance of total prima facie rightness over total prima facie wrongness. The total prima facie rightness of an action is simply the sum of the strengths of all of the prima facie duties the agent has *to* perform it. The total prima facie wrongness of an action is the sum of the strengths of all of the prima facie duties the agent has *not* to perform it. Prima facie

¹¹ For more on how to understand what it is to treat humanity as an end in itself, see Hill (1980), Wood (1998), Pogge (1998), and Dean (2009).

duties, then, have strengths. And the idea is that for any action a person can perform, we can look at all of the prima facie duties she has to do it and add up their strengths, and we can look at all of the prima facie duties she has *not* to do it and add up their strengths. Then we can subtract the latter sum from the former. If the resulting value is higher than that of any other action the person can perform instead, then it is obligatory.

Of course, stated this way, Ross's theory is very abstract. It's also incomplete. We still need to be told what prima facie duties we have, and we need to be told how to calculate their strengths. With regard to the first issue, Ross says a great deal, as we've seen. However, Ross says very little about the latter issue. In *R&G*, he says virtually nothing about how to calculate the strengths of prima facie duties. But in *FE*, he does address the issue briefly, at least with regard to the prima facie duty of promise keeping. There he says that the strength of the duty to keep a promise is, in part, a function of its "importance" to the promisee (1939, 100). The idea here seems to be this. If I make a promise to you and keeping it would please you greatly, then, other things being equal, I have a stronger prima facie duty to keep it than I would have if keeping the promise would only please you slightly. Ross also thinks that the strength of the duty to keep a promise is partly a function of its solemnity (100-1). In other words, the prima facie duty to keep a solemn promise (perhaps made under oath) is, other things being equal, stronger than the prima facie duty to keep a casual or offhand promise. Finally, Ross suggests that the strength of the duty to keep a promise is, in part, a function of its recency: other things being equal, the duty to keep a promise that has just been made is stronger than the duty to keep a promise that was made long ago (101).

This, of course, is merely a sketch. More needs to be said before we can calculate the strength of the prima facie duty of an actual promise keeping.¹² But for my purposes, the important point is that Ross seems to think that the strength of the duty of promise keeping is determined by purely descriptive facts—namely, by the promise’s recency, its solemnity, and its importance to the promisee. And while Ross never explicitly says so in his work, I suspect that he’d say that the strengths of the other prima facie duties are, at bottom, determined by descriptive facts too. But if he thinks this, then, since he thinks that moral obligation is determined by the strengths of prima duties, he would presumably also admit that obligation is, at bottom, determined by descriptive facts. He would, in other words, admit that there is a true absolute moral principle as defined by MP4.

I realize that I’m going somewhat beyond the text here. Ross never explicitly accepts a proposition that counts as an absolute moral principle according to MP4. But he clearly does think that obligation depends on the strengths of prima facie duties, and he does suggest, at least with regard to promise keeping, that the strengths of prima facie duties depend on purely descriptive facts. So I think it’s plausible to hold that Ross would, if pushed, admit that there is a true proposition that has the features necessary for being an absolute moral principle on MP4. Of course, such a proposition would be long and complicated.¹³ Ross would likely say that we don’t know, or perhaps can’t know,

¹² For instance, while Ross tells us which things the strength of the duty of promise keeping is a function of, he never indicates precisely what this function is.

¹³ Though it wouldn’t be nearly as long and complicated as the one generated by the supervenience of the moral on the natural, for its right-hand side would not mention *all* the

exactly what it looks like. But I see no reason to doubt that Ross would accept that such a proposition exists.

So, I think that MP4 is not a conception of absolute moral principles that has the features we're looking for. It doesn't commit all utilitarians—or even the most important ones—to generalism about absolute duty. It may not commit all Kantians to generalism about it either. Moreover, there is reason to think that it doesn't even commit Ross to particularism about this sort of duty.¹⁴

descriptive facts of a world, but rather, only those descriptive facts that determine the stringencies of prima facie duties.

¹⁴ Someone might wonder whether the claims that utilitarians and Kantians accept constitute a much “deeper” or “robust” explanation of deontic concepts than do the claims that Ross accepts (in particular, RTO_{RG}). If utilitarians and Kantians do indeed accept deeper explanations of deontic concepts than Ross does, perhaps this is what makes utilitarians and Kantians generalists about absolute duty and Ross a particularist about it. However, if I am right that Ross would accept that there is a true absolute moral principle as defined by MP4, then he would seem to be on a par with many utilitarians and Kantians vis-à-vis explanatory depth. Consider for a moment the views of one particular utilitarian: G. E. Moore. As we've seen, he holds that an action is obligatory iff (and because) it maximizes the overall balance of intrinsic goodness. Moore ([1903] 1993) also adopts a pluralistic theory of the good: he suggests that beauty, the appreciation of beauty, personal affection, and pleasure (taken in good things) are all intrinsically good. In addition, Moore would surely admit that the degree to which a beautiful object (a feeling of personal affection, etc.) is intrinsically good is determined by further descriptive facts. Now compare this package of claims with ones that Ross accepts. Ross holds that an action is obligatory iff (and because) it maximizes the overall balance of prima facie duty. Ross also accepts a pluralistic theory of prima facie duty: he thinks that beneficence, reparation, justice, etc. are all prima facie duties. Moreover, as I've suggested, there is no reason to doubt that Ross would say that the degree to which an act of beneficence (reparation, etc.) is a prima facie duty is determined by further descriptive facts. There is thus a striking symmetry between the claims that Ross and Moore accept. Given this, it is difficult to see one as offering a more robust explanation of moral obligation than the other. Similar things will apply, I think, when we compare the claims that Ross accepts with the claims of other utilitarians and Kantians. The general point, then, is that if Ross accepts that there is a true moral principle as defined by MP4,

2.5. What's an Absolute Moral Principle?—Some Final Answers

I have suggested that Ross may in fact accept that there is a true proposition that has the features needed to make it an absolute moral principle according to MP4. However, I admitted that such a principle would likely be long and complicated. It therefore wouldn't be useful as a guide to action. In other words, it would be very difficult for us to use it to determine whether our actions are obligatory. But some particularists suggest that moral principles are, by their very nature, useful, or action-guiding.¹⁵ In light of this, they might say that MP4 needs to be modified as follows:

MP5: An absolute moral principle is, by definition, a non-trivial, **action-guiding** proposition that has the following form: an action is morally obligatory (permissible, wrong) iff and in virtue of the fact that _____, where this blank contains only descriptive terms.

Will Ross count as a particularist about absolute duty on this conception of moral principles? I suspect that the answer is “yes.”¹⁶ However, it seems to me that many utilitarians will also be particularists about absolute duty on MP5. Consider, for instance, hedonistic act utilitarianism. This theory is rarely action-guiding; we can almost never

then no matter how we cash out the notion of explanatory depth, it looks as though the explanations that Ross endorses will be no less deep than ones accepted by many generalists.

¹⁵ See Dancy (2004b, 87-88; 1993, 69, 77-79), Little (2000), and Lance and Little (2006b, 570-73). In chapter 7, I will say more about precisely what it means for a moral theory to be “action-guiding.” For now, I will simply assume that this idea is well-enough understood.

¹⁶ That Ross would deny that there are any true absolute moral principles that are action-guiding is, perhaps, suggested by his (1930, 31).

use it to determine the moral status of our actions because we almost never know the precise amount of pleasure or pain our alternatives will bring about.¹⁷

Of course, there are other ways of formulating utilitarianism, but most of them don't make the theory any more action-guiding. Consider, for instance, Moore's Ideal Utilitarianism. It is no easier to determine how much intrinsic goodness and intrinsic badness our alternatives will produce than it is to determine how much pleasure and pain they will produce. Ideal Utilitarianism thus seems to be an extremely poor guide to action, as Moore ([1903] 1993, secs. 91-94, 99) seems to realize.

Others formulate utilitarianism in a different way. They take it to be the view that an action is obligatory iff its expected utility is greater than the expected utility of any of its alternatives. However, as some have pointed out (especially Feldman 2006), this version of utilitarianism seems even less action-guiding than traditional act utilitarianism. After all, the expected utility of an alternative is the sum, for each of its possible outcomes, of the outcome's actual value times the probability that it (the outcome) will obtain if the alternative is performed. More precisely, where A is an alternative, O_1, O_2, \dots, O_n are the possible outcomes of A , $V(O)$ is the value of outcome O , and $\text{prob}(O/A)$ is the probability of O given A , the expected utility of $A = \{[\text{prob}(O_1/A) \times V(O_1)] + [\text{prob}(O_2/A) \times V(O_2)] + \dots + [\text{prob}(O_n/A) \times V(O_n)]\}$. Thus, to determine the expected utility of just one of our alternatives, we would need to determine not only the actual value of each of its possible outcomes, but we'd also need to determine the probability that each of these outcomes will obtain if the action is performed. But these are things

¹⁷ See Ross (1930, 23-24).

that we will almost never be able to determine, at least in any realistic scenario. And even if we could determine these things, we'd still have a multitude of mathematical calculations to perform—we'd still have to multiply the value of each of the alternative's possible outcomes by the probability that it will obtain if the alternative is performed. Such a task is one that we could almost never carry out. And even if we could, in order to determine whether the alternative in question is obligatory according to expected utility utilitarianism, we'd still have to repeat this whole process for each of our other alternatives and compare their expected utilities—a daunting task indeed.

It thus seems that many of the main formulations of utilitarianism fail to provide us with absolute moral principles as defined by MP5.¹⁸ I suspect that similar things can be said about many formulations of Kantianism. Consider, for instance, the first formulation of the Categorical Imperative, according to which an action is morally permissible iff its maxim is universalizable. In order for an agent to use this theory to obtain moral guidance, she'll presumably need to be able to determine (a) the maxims associated with her actions, and (b) whether these maxims are universalizable. However, I suspect that on many ways of understanding what a maxim is, it will be possible (perhaps even common) for an agent to have difficulty determining what the maxims of her actions are. For instance, if we think of the maxim of an action as the motive or intention that the action is performed on, then we will surely sometimes be unsure what the maxims of our actions are, for we are surely sometimes unsure what motives or intentions we act upon. Similar things seem to apply if we think of the maxim of an action as the underlying

¹⁸ For more on action-guidingness and utilitarianism/consequentialism, see Frazier (1994), Lenman (2000), and Miller (2003).

policy, aim, or principle that the agent acts on. And even if we could always identify the maxims of our actions, we may not always be able to determine whether they are universalizable. Suppose, for instance, that for the maxim of an action to be universalizable is for the action to be logically possible in a world where everyone acts on it (the maxim). However, it is difficult, in many circumstances, to determine precisely what the world would be like if everyone in it acted on a given maxim. So, I suspect that on this understanding of universalizability, agents will sometimes (perhaps often) be unsure about whether the maxims of their actions are universalizable. Similar things also seem to apply if we hold instead that for the maxim of an action to be universalizable is for the agent to be able to achieve the end of her action even in a world where everyone acts on the action's maxim.¹⁹ Now, I don't take any of this to definitively show that Kantianism, in all of its forms, fails to be action-guiding. However, I think it at least shows that a *prima facie* case can be made that Kantianism, in at least some of its forms, is not fully action-guiding. And if I have successfully shown this, then I have also shown that a *prima facie* case can be made that MP5 will deem many Kantians—as it will many utilitarians—to be particularists about absolute duty.

Before concluding, I'd like to briefly consider one final account of absolute moral principles. Some generalists think that principles that have many of the features specified by MP5 are “constitutive” of moral judgment. That is, they think that in order to have the concepts of moral obligation, permissibility, and wrongness, one needs to accept

¹⁹ The two ways of understanding universalizability discussed here correspond roughly to what has become known as the “Logical Contradiction Interpretation” and the “Practical Contradiction Interpretation,” respectively (see Korsgaard 1996, 78).

principles of the sort suggested by MP5.²⁰ This may suggest another way of conceiving of absolute moral principles. Perhaps absolute moral principles don't merely specify what moral obligation (permissibility, wrongness) is coinstantiated with, nor do they tell us what grounds moral obligation (permissibility, wrongness), perhaps they instead provide us with an *analysis* of the concept of moral obligation (permissibility, wrongness). In other words, maybe we should understand absolute moral principles more along the lines of this:

MP6: An absolute moral principle is, by definition, a non-trivial proposition that has the following form: an action is morally obligatory (permissible, wrong) =*df.* _____, where this blank contains only descriptive terms.

If we accept MP6, I think we should again conclude that Ross is a particularist about absolute duty. He clearly believes that we can't give a descriptive analysis or definition of "obligation" (1930, chap. 1; 1939, chap. 2). So, he'd deny that there are any true absolute moral principles as understood by MP6. However, if we accept MP6, it again seems that we'll also have to admit that many utilitarians are particularists about absolute duty. While some of the early utilitarians, such as Mill and Bentham, may have thought that utilitarianism provides us with an analysis or definition of "obligation," utilitarianism isn't typically understood in this way today.²¹ Moreover, many important

²⁰ Jackson, Petit, and Smith (2000) defend constitutive generalism. For discussion of their view, see McKeever and Ridge (2006, chap. 5).

²¹ Utilitarianism is now typically understood as being a theory about the necessary and sufficient conditions for the moral obligatoriness (permissibility, wrongness) of actions, or a theory about the feature in virtue of which all obligatory (permissible, wrong) actions are obligatory (permissible, wrong). See, for instance, Feldman (1978, 26), Timmons (2002, 104-6), and Shafer-Landau (2012, 119-20).

utilitarians—such as Sidgwick (1907) and Moore ([1903] 1993)—explicitly deny that a descriptive analysis of deontic terms such as “obligation” can be given.²² So, I think that if we accept MP6, we’ll have to admit that while Ross would likely deny that there are any true absolute moral principles, so would many utilitarians.

2.6. Conclusion and Application to Dancy

I have now considered several ways of understanding what an absolute moral principle is, each of which is derived from things that actual particularists and generalists say about moral principles. I have shown that on many of these conceptions (MP1-MP4), Ross seems to *accept* absolute moral principles (or, at least, we have no good reason to think otherwise). On these conceptions, Ross is therefore not a particularist about absolute duty, as he is widely seen to be. However, I also considered several ways of conceiving of absolute moral principles (MP5 and MP6) according to which Ross does seem to be a particularist about this kind of duty. But I argued that on these conceptions, many utilitarians and Kantians will also be particularists about it, not generalists, as they are widely assumed to be. My conclusion, then, is that Ross is no more a particularist about absolute duty than a utilitarian or a Kantian is. When it comes to affirming the existence of true absolute moral principles, Ross belongs in the same camp as utilitarians and Kantians. If we want to make utilitarians and Kantians generalists about absolute duty, that’s fine, but we should then admit that Ross is a generalist about it too. If we’d like to make Ross a particularist about this sort of duty, that’s fine too, but we should

²² Of course, Moore ([1903] 1993) claims that a *non*-natural analysis of obligation can be given: he analyzes obligation in terms of goodness and badness. However, he denies that a purely descriptive analysis of goodness and badness can be given.

then hold that utilitarians and Kantians are particularists about it as well. But we can't have it both ways; there doesn't seem to be a plausible conception of absolute moral principles that will make Ross a particularist about absolute duty and utilitarians and Kantians generalists about it.

I think this conclusion is important. Too often Ross is considered to be fundamentally at odds with utilitarians and Kantians. Of course, Ross disagrees heartily with utilitarians and Kantians about many things. But as I see it, Ross is engaged in the same overarching project as they are. Ross, like utilitarians and Kantians, is attempting to provide a fundamental theory of moral obligation. That is, he is trying to give a non-trivial, explanatory set of necessary and sufficient conditions for the moral obligatoriness of actions. And the theory of obligation that Ross offers (namely RTO_{RG}) is, I think, both interesting and important. It deserves to be treated alongside the theories that utilitarians and Kantians offer. However, if we think that Ross is a particularist about absolute duty while utilitarians and Kantians are generalists about it, we run the risk of neglecting Ross's theory in favor of the utilitarian and Kantian theories. After all, a fundamental theory of obligation certainly looks like an absolute moral principle, at least at first glance. So, if we hold that Ross denies while utilitarians and Kantians accept absolute moral principles, it's all too easy to ignore Ross's theory altogether and consider only utilitarian and Kantian theories. And that, I think, is something we should be careful to avoid.

My conclusion is also important for another reason, for it may have implications on the broader debate between particularists and generalists. This is because I suspect that many self-proclaimed particularists would accept much of what I've argued that Ross

accepts. That is, I suspect that many particularists would accept a proposition like RTO_{RG} (or some minor variant of it), deny that it is action-guiding, deny that it is a conceptual truth, etc. But if particularists accept these things, then I would argue that, like Ross, these “particularists” are no more particularists about absolute duty than many utilitarians or Kantians are. If a particularist wants to resist this conclusion, she’ll need to provide us with a conception of absolute moral principles that will render her a particularist about absolute duty without also rendering utilitarians or Kantians particularists about it. Given what I’ve said in this chapter, it should be clear that I suspect she’ll have a hard time succeeding.

It would be especially interesting to see whether Jonathan Dancy—perhaps the most well-known particularist—can rise to the challenge. Detailed discussion of Dancy’s rich and complex moral outlook will have to await another occasion. However, I will close this chapter by briefly considering some of his views. First of all, Dancy (2004b, chap. 5) accepts what he calls “holism,” both about contributory reasons and about ought-making features. According to holism about contributory reasons (i.e., reasons that count in favor of an action), a feature that is a reason for an action in one context might fail to be a reason for an action in another context (or it might even be a reason *against* the action). According to holism about ought-making features, a feature that makes an action obligatory in one context might fail to make an action obligatory in another context.²³ (Dancy says similar things about permissible-making features as well as wrong-making features.) Because of Dancy’s commitment to holism, he might simply accept a

²³ It’s unclear to me whether Dancy would say that it’s possible for a feature to make an action obligatory in one context but *wrong* in another.

conception of moral principles like MP2, according to which an absolute moral principle is a proposition of the form “an action is morally obligatory (permissible, wrong) iff and in virtue of the fact that _____.” He might then go on to claim that there are no true propositions of this sort: the blank simply cannot be filled in because ought- (permissible-, wrong-) making features are holistic. Utilitarians and Kantians, however, clearly do think that there are true absolute moral principles that conform to MP2 (as does Ross, I’ve argued). So, Dancy may in fact be able to rise to the challenge I’ve raised.

However, the matter is complicated by the fact that Dancy (2004b, 29-37) appears to accept something along the lines of Ross’s R1, though Dancy prefers to state it in terms of “contributory oughts” instead of in terms of “prima facie duties.”²⁴ It’s unclear to me precisely what Dancy has in mind by a “contributory ought”; however, he ultimately appears to understand “contributory oughts” in terms of “peremptory reasons.” (Recall that for Dancy, a peremptory reason just is what I earlier called a “requiring reason.”²⁵) The upshot of all of this is that I suspect that Dancy would accept something along the lines of the claim that an action is obligatory iff there is more overall peremptory reason to perform it than there is to perform any of its alternatives. But if Dancy would indeed accept this (or something like it), it’d be natural to wonder: would he also admit (as Ross clearly does about R1) that obligatory actions are obligatory *in virtue of the fact* that there is most overall peremptory reason to perform them? It’s

²⁴ He says, for instance, that “I have been gently moving towards the conclusion that the overall ought should be understood as some function of a contributory ought” (2004b, 34).

²⁵ See note 8.

unclear to me what Dancy would say about this. But if he'd say "yes," then it looks like he'd accept absolute moral principles according to MP2 after all.

At this point, Dancy might counter by saying that even if obligatory actions are obligatory in virtue of maximizing the overall balance of preemptory reasons, this is not enough to make him a generalist about absolute duty because such a "principle," even if it is to some extent explanatory, is certainly not action-guiding. Such a response would suggest that Dancy accepts a conception of absolute moral principles more like MP5. And indeed, there are several passages in *Ethics without Principles* where Dancy seems to adopt something like this conception of moral principles (see, for instance, his discussion of resultance and supervenience [2004b, 85-93]). But, as I've argued, even many utilitarians and Kantians will deny that there are true principles of this sort, and so, if Dancy accepts a conception of principles like this, he may not be able to rise to the challenge after all.

Of course, even if there is no good way of understanding what absolute moral principles are that will make Dancy a particularist about them and utilitarians and Kantians generalists about them, it might still be the case that Dancy is a particularist while utilitarians and Kantians are generalists *on some other sense* of the terms "particularism" and "generalism." As I noted above, there are lots of ways of understanding "particularism" and "generalism"; according to the way that I have been using the terms in this chapter, they refer to the acceptance (or denial) of true moral principles.²⁶ However, it's worth noting that Dancy now seems to understand these terms

²⁶ See note 1.

in a quite different way. For him, particularism is the view that “the possibility of moral thought and judgement does not depend on the provision of a suitable supply of moral principles” (2004b, 7, 73). Generalism, on the other hand, is the view that moral thought and judgment does indeed depend on the provision of a suitable supply of moral principles. But to claim that the possibility of moral thought and judgment depends (or doesn’t depend) on moral principles seems very different from the mere claim that there are some (or are no) true moral principles. Nothing I’ve said here has shown that Dancy is a generalist on his idiosyncratic sense of the term. Nor have I done anything to show that utilitarians and Kantians are particularists on Dancy’s sense of the term. (Though I think a case can be made, at least for the latter claim, but I won’t press the point here.)

Finally, let me stress that if it turns out that my argument implies that Dancy is no more a particularist about absolute duty than a utilitarian or a Kantian is, I wouldn’t take this to be an objection to anything I’ve said. Indeed, while my main concern has been to make this point about Ross, not Dancy, one of the undercurrents of the chapter has been that what I say about Ross may very well apply to other paradigmatic particularists. If it does indeed apply even to Dancy, this would merely confirm my suspicion. Though it bears repeating that Dancy’s views are complicated, and what I’ve said about them has been somewhat speculative. The extent to which what I’ve said about Ross applies also to Dancy (or to any other paradigmatic particularist) is a topic that deserves more consideration than I am able to give it here. I mention it merely as an illustration of what such an application might look like, or, at least, how it might begin.

CHAPTER 3

ROSS AND PLURALISM

3.1. Introduction

While many philosophers claim that Ross is a type of particularist, even more claim that he is a pluralist.¹ Philosophers who suggest that Ross is a pluralist include Stratton-Lake (2000, 2002, 2011b), McNaughton (1988, 1996, 2000), Dancy (1983, 1991a, 1991b), Audi (1996, 2004), Rawls (1999), Hooker (1996, 2000), Timmons (2002), Shafer-Landau (2012), Gibbard (2003), Gaut (2002), Berker (2007), Montague (2000), McCann (2007), and McConnell (1988).² In short, virtually every philosopher who has written about Ross suggests that he's a pluralist. However, in this chapter, I will challenge this widespread assumption.

Here's my plan for the chapter. In order to determine whether Ross is a pluralist, we first need to determine what pluralism is. However, this is not an easy thing to determine, for there are *many* different things that philosophers mean by the term "pluralism." There are, in short, many distinct doctrines of pluralism. In what follows, I will identify several conceptions of pluralism present in the literature, and I will argue

¹ That is, many philosophers suggest that Ross is a *deontic* pluralist. Deontic pluralism is pluralism about permissibility, wrongness, and obligation. Many philosophers also suggest that Ross is a *value* pluralist. Value pluralism is pluralism about intrinsic goodness and badness. In this chapter, I'll be discussing whether Ross is a deontic pluralist, not whether he is a value pluralist. By "pluralism," I'll always mean "deontic pluralism."

² I will quote specific passages from most of these sources in the course of this chapter.

that on each one, either Ross is clearly not a pluralist, or there is no reason to think that he is.³

3.2. Pluralism(c)

According to perhaps the most common way of understanding pluralism, it is the view that there is more than one feature in virtue of which an action can be overall morally obligatory/permissible/wrong.⁴ Equivalently, it's the claim that there's more than one property the possession of which can make an action overall morally obligatory/permissible/wrong. I will call this conception of pluralism "pluralism(c)" ("c" for "common"). In light of this, we can give a corresponding conception of monism (pluralism's arch-nemesis). Monism(c) is the view that there is only *one* feature in virtue of which an action can be overall morally obligatory/permissible/wrong. Equivalently, it's the view that there's only one property the possession of which can make an action overall morally obligatory/permissible/wrong.

Theories like act utilitarianism, Kantianism, contractualism, and the divine command theory are supposed to be paradigmatic examples of monistic(c) theories. Act utilitarians, for instance, hold that the property of maximizing utility is the only feature that can make an action overall morally obligatory. Kantians hold that the property of treating humanity as an end in itself, not as a mere means, is the only overall obligatory-

³ I am again going to confine my attention to the Ross of *R&G*. However, I think similar things can be said of the Ross of *FE*.

⁴ Below, I will cite several examples of authors who understand pluralism in this way.

making feature.⁵ And divine command theorists hold that the property of conforming to God's commands is the only feature in virtue of which an action can be overall morally obligatory.

However, some philosophers recently have rejected all forms of monism(c) and defended various forms of pluralism(c) instead. Dancy (2004b) and Stratton-Lake (2000, 2011b) are two prominent examples. These philosophers often take inspiration from Ross. Consider, for instance, what Philip Stratton-Lake says in the following passages:

As I understand prima facie duties they are recommenders not recommendations. They do not tell you what you should do, but pick out features that give you reason to do certain acts. This point is quite lost in the terminology Ross uses in *The Right and the Good*. But despite the misleading terminology, it is clear that by "prima facie duty" he does not mean a special sort of duty, or recommendation. He means, rather, a moral feature of acts related to duty proper. The context in which Ross introduces the term, as well as much of what he says about the notion of a prima facie duty, makes it clear that the way in which he thinks that prima facie duties are related to duties proper is by explaining them. Principles of prima facie duty pick out features of acts or situations that tend to make certain acts right. So if they are not defeated in some way they will pick out the reason why some act is right. This means that principles of prima facie duty pick out explanatory reasons—the *reason why* some act is our duty, or the *reason why* we should do that act, or *why* it is right. (2011b, 367-68)

Ross maintains that the reason why I ought to keep my promise to A to do a certain act is simply that I have promised A that I would do that act; and the reason why I ought to be grateful to B is that B benefited me in the past. These reasons are captured in the principles of fidelity and gratitude. The principle of fidelity states that the fact that I have promised A to Φ gives me a moral reason to Φ , and the principle of gratitude states that the fact that B did me a favour in the past gives me a reason to

⁵ Some Kantians may perhaps hold instead that some other Kantian-type property (such as the property of having a universalizable maxim, or the property of conforming to the maxims of a member giving universal laws for a merely possible kingdom of ends) is the obligatory-making feature.

be grateful to B. Assuming that these reasons are not defeated, it will turn out that my duty proper is to keep my promise, or be grateful, and the facts mentioned in these principles will explain these duties. Since it is the fact that I promised A that explains my obligation to do what I have promised, it is A who is wronged if I fail to keep my promise. Since it is the fact that B benefited me in the past that explains my obligation to be grateful to B, it is B who is wronged if I miss an opportunity to express my gratitude to B, or reciprocate. (2011b, 375)

In these passages, Stratton-Lake attributes to Ross the following type of view. When we have an absolute duty to keep a promise, we have such a duty because of the fact that the action fulfills a promise. On the other hand, when we have an absolute duty to perform an act of gratitude, what explains why we have this duty is the fact that the action is an act of gratitude. And so forth, Stratton-Lake would presumably say, for acts of reparation, justice, beneficence, etc. Stratton-Lake therefore appears to attribute pluralism(c) to Ross: on his view, Ross holds that many different features can make an action overall morally obligatory, not just one.

Stratton-Lake is by no means the only philosophers to interpret Ross in this manner.

Alan Gibbard, for instance, says the following about Ross:

Normative pluralists maintain the thesis, sometimes, that no single property always explains why to do those things we ought to do. Sometimes we ought to do something because it's keeping a promise at a small loss in general happiness, sometimes because it's crucial to one's self-development and requires breaking only a minor promise—and so on. Perhaps some such way as this is the best way to put the views of a pluralist like Ross. (2003, 151)

Hugh McCann says something similar:

Moral intuitionism fits well, then, with the distinctly nontheoretical tone of much of everyday ethical life. But it hardly follows that an intuitionist ought to eschew ethical theory. Ross himself did so, considering *prima facie* duties to be self-evident but not

provable, and denying that there was any one general characteristic that makes right acts right. (2007, 42)⁶

But are these philosophers right to attribute pluralism(c) to Ross? There are some passages that may seem to suggest that they are. Consider, for instance, what Ross says here:

When a plain man fulfils a promise because he thinks he ought to do so, it seems clear that he does so with no thought of its total consequences, still less with any opinion that these are likely to be the best possible. He thinks in fact much more of the past than of the future. What makes him think it right to act in a certain way is the fact that he has promised to do so—that and, usually, nothing more. That his act will produce the best consequences is not his reason for calling it right. (1930, 17)

In this passage, Ross suggests that when “plain men” think they ought to keep a promise, they think they ought to do so simply because doing so will fulfill a promise, not because it, say, has the best consequences. But if Ross thinks this, then he would surely also say similar things about acts of reparation, justice, gratitude, etc. That is, he would say that when plain men think they ought to perform an act of reparation (justice, gratitude, etc.), they think they ought to do so simply because doing so will be an act of reparation (justice, gratitude, etc.). The above passage (hereafter “the plain man passage”) therefore seems to suggest that Ross believes that ordinary people are pluralists(c): they hold that many different features can make an action morally obligatory.

To be sure, in the plain man passage, Ross doesn’t explicitly say that *he* accepts pluralism(c). At most, he suggests that the *ordinary person* (or the “plain man”) accepts

⁶ See also Robinson (2010, 26).

it. However, throughout *R&G*, Ross takes the judgments of ordinary people very seriously. He says,

The main moral convictions of the plain man seem to me to be, not opinions which it is for philosophy to prove or disprove, but knowledge from the start. (1930, 20-21n)

Later on, he says,

I would maintain, in fact, that what we are apt to describe as “what we think” about moral questions contains a considerable amount that we do not think but know, and that this forms the standard by reference to which the truth of any moral theory has to be tested, instead of having itself to be tested by reference to any theory. I hope that I have in what precedes indicated what in my view these elements of knowledge are that are involved in our ordinary moral consciousness. (1930, 40)

So, although Ross only suggests in the plain man passage that ordinary people accept pluralism, one might think that there is good reason to believe that he accepts it as well, or, at least, is highly sympathetic to it.

However, it's important to realize that the plain man passage comes toward the beginning of chapter 2 of *R&G*. I think Ross is speaking a bit loosely in the passage. Indeed, Ross seems to acknowledge this later on in the chapter. Toward the end of the chapter, Ross points out that actions are describable in many ways. Suppose, for instance, that I have promised to return a book to a friend. After finishing the book, I pack it into a suitable box, affix the correct postage, and drop it into an appropriate mailbox. After that, the box is handled by various employees at the post office and is eventually delivered to my friend. Ross claims that in this case, I have done my duty, but my action can be described in a multitude of ways. For instance, it can be described as the packing and posting of the book, the provision of a few moments of employment for post

office officials, the fulfillment of my promise to my friend, etc. Given this, one might wonder: in virtue of which of these descriptions have I done my duty? Ross claims that it's in virtue of the third of these descriptions. As he says,

And if we ask ourselves whether it is *qua* the packing and posting of a book, or *qua* the securing of my friend's getting what I have promised to return to him, that my action is right, it is clear that it is in the second capacity that it is right; and in this capacity, the only capacity in which it is right, it is right by its own nature and not because of its consequences. (1930, 44)

Shortly after saying this, however, Ross qualifies his claim slightly. He says (and this is the important part of the passage for our purposes, a passage we also encountered in chapter 1 (sec. 1.4)):

We have reached the result that my act is right *qua* being an ensuring of one of the particular states of affairs of which it is an ensuring, viz., in the case we have taken, of my friend's receiving the book I have promised to return to him. But this answer requires some correction; for it refers only to the *prima facie* rightness of my act. If to be a fulfillment of promise were a sufficient ground of the rightness of an act, all fulfillments of promises would be right, whereas it seems clear that there are cases in which some other *prima facie* duty overrides the *prima facie* duty of fulfilling a promise. The more correct answer would be that the ground of the actual rightness of the act is that, of all acts possible for the agent in the circumstances, it is that whose *prima facie* rightness in the respects in which it is *prima facie* right most outweighs its *prima facie* wrongness in any respects in which it is *prima facie* wrong. But since its *prima facie* rightness is mainly due to its being a fulfillment of promise, we may call its being so the salient element in the ground of its rightness. (1930, 45-46)

Here Ross admits that when he suggested that an action is right (i.e., obligatory) in virtue of being the fulfillment of a promise, what he really meant was that the action is *prima facie* right in virtue of being the fulfillment of a promise. When we turn to absolute duty, however, things are different. An action is never an absolute duty merely because it fulfills a promise. Rather, an action is overall obligatory in virtue of having a greater

total balance of prima facie rightness over wrongness than any alternatives does. And, Ross suggests in this and other passages (see sec. 1.4), all obligatory actions are obligatory in virtue of having this feature. Ross, then, looks to be a monist(c), not a pluralist(c); contrary to what Stratton-Lake and others suggest, Ross appears to believe that there is only one feature that makes actions overall obligatory, not many such features.

3.3. Pluralism(n&s)

A conception of pluralism similar to pluralism(c) is suggested by Phillip Montague and can be dealt with quickly. Montague says:

In its standard form, act consequentialism can be considered monistic, since it implies that an action's possessing a certain single feature (that of maximizing aggregate value) is necessary and sufficient for that action to be morally required. Thought of in these terms, rule consequentialism is pluralistic—as are deontological theories of the sort espoused by W. D. Ross. (2000, 203)

Montague suggests here that monism is the view that there is a feature that's necessary and sufficient for an action to be (overall) morally obligatory. Pluralism, then, is presumably the view that there is no feature that's necessary and sufficient for an action to be (overall) morally obligatory. Let's call this form of monism and pluralism, "monism(n&s)" and "pluralism(n&s)," respectively ("n&s" for "necessary and sufficient conditions").⁷

⁷ Other philosophers suggest a similar conception of pluralism. For instance, Terrance McConnell says that Ross subscribes to "irreducible pluralism," which is "the view that there is nothing that all and only right actions have in common except their rightness" (1988, 91).

Montague further claims that while consequentialists are monists(n&s), Ross is a pluralist(n&s). However, as we've seen, Montague is plainly wrong about this. In virtue of his acceptance of RTO_{RG} , Ross holds that there *is* a feature that's necessary and sufficient for an action to be (overall) morally obligatory. He is therefore not a pluralist(n&s).

Notice also that virtually all moral theorists will admit, as Ross does, that there's some feature that's necessary and sufficient for an action to be (overall) morally obligatory. Everyone will presumably agree, for instance, that an action is obligatory iff it is wrong not to perform, and that an action is obligatory iff it is permissible and no alternative to it is permissible. Consider also a giant list of all actions that could be obligatory. Call this list "L." All will agree that an action is obligatory iff it is on list L. Or consider a giant list of the natural supervenience bases of all actions that could be obligatory. Call this list "B." All moral theorists—or, at least, all those who accept that the moral supervenes on the natural—will agree that an action is obligatory iff its supervenience base is on list B. The point is that it's very easy to find necessary and sufficient conditions for the moral obligatoriness of actions. Because of this, not only will Ross qualify as a monist on monism(n&s), but so will virtually every other moral theorist.⁸

⁸ There is another way of interpreting Montague in the above passage. Perhaps his idea is that a monist is someone who thinks that there's a *single* feature (i.e., exactly one feature) that's necessary and sufficient for an action to be morally obligatory. Pluralism, then, is the view that there's *more than one* feature that's necessary and sufficient for an action to be morally obligatory. However, given what I've said in the paragraph this footnote is appended to, it should be clear that virtually every moral theorist, Ross included, will accept that there are multiple

3.4. Pluralism(3)

According to another popular way of understanding pluralism, it has three tenets. The first tenet claims that there are several fundamental *prima facie* duties. The second tenet claims that these fundamental *prima facie* duties can conflict in certain circumstances—in other words, there can be actions that the agent has both a fundamental *prima facie* duty to perform and a fundamental *prima facie* duty not to perform. The third tenet claims that in cases where there is a conflict of fundamental *prima facie* duties, there exists no finite principle that can, in each of these cases, resolve the conflict—more precisely, there is no finite principle that can tell us, for each possible action that the agent has both a fundamental *prima facie* duty to perform and a fundamental *prima facie* duty not to perform, which duty is stronger. I'll call this version of pluralism “pluralism(3)” (“3” for “3 Tenets”). Monism(3), then, is the view that at least one of the three tenets of pluralism(3) is false.

Several philosophers suggest that Ross is a pluralist(3). Selim Berker is one such philosopher. Consider what he says here:

Among the philosophers sympathetic to this sort of criticism of the monist tradition, some have chosen to embrace a *pluralist* approach to ethics such as that found in W. D. Ross's theory of *prima facie* duties or in some (but not all) forms of virtue ethics, where instead of one solitary moral principle there is posited to be a plurality of basic principles—an “unconnected heap of duties,” to use David McNaughton's apt expression, that are all equally fundamental. These differing duties can in principle conflict, but in such cases there are claimed to be no finitely codifiable rules dictating which duty trumps or outweighs the others, for otherwise the basic principles together with the weighing rules could be conjoined into a single master principle. (2007, 110)

features necessary and sufficient for the moral obligatoriness of actions, and that virtually all moral theorists will be pluralists on this understanding of the pluralism.

Likewise, Berys Gaut claims that Ross adopted moral pluralism (2002), which he understands as follows:

Moral pluralism is the theory that there is a plurality of first-order moral principles stating what one has moral reason to do [i.e., *prima facie* duties]; that these principles may conflict in their application to particular cases; and that there is no higher-order moral principle which in each case of conflict ranks one first-order principle over another. What one has to do in such cases is to employ one's judgment about what one has most reason to do, a judgment which escapes codification by principles. (138)

However, it seems to me that Berker, Gaut, and others,⁹ are wrong to think that Ross is a pluralist(3). Or, at least, there is no good reason to think that Ross accepts it.

Ross obviously accepts the first tenet of pluralism(3); he clearly thinks there are multiple fundamental *prima facie* duties. He also surely accepts the second tenet. He thinks that our fundamental *prima facie* duties often conflict. He says, "moral acts often (as every one knows) and indeed always (as on reflection we must admit) have different characteristics that tend to make them at the same time *prima facie* right and *prima facie* wrong; there is probably no act, for instance, which does good to any one without doing harm to some one else, and *vice versa*" (1930, 33-34; see also 23, 30, 41). However, while Ross accepts the first two tenets of pluralism(3), I see no reason to believe that he accepts the third.

⁹ John Rawls may also attribute this type of pluralism to Ross. He suggests that Ross is an intuitionist/pluralist (1999, 30n), and he seems to understand intuitionism/pluralism as pluralism(3). He says, "Intuitionist theories, then, have two features: first, they consist of a plurality of first principles which may conflict to give contrary directives in particular types of cases; and second, they include no explicit method, no priority rules, for weighing these principles against one another: we are simply to strike a balance by intuition, by what seems to us most nearly right....Perhaps it would be better if we were to speak of intuitionism in this broad sense as pluralism" (1999, 30-31). See also, DePaul (1988, 73), McNaughton (1988, 197-98), and Lance and Little (2008, 55).

As I've interpreted the third tenet of pluralism(3), it's the claim that there is no finite principle that can tell us, for each possible action that the agent has both a fundamental prima facie duty to perform and a fundamental prima facie duty not to perform, which duty is stronger. But consider this principle: in any case where an agent has an option that she has both a fundamental prima facie duty to perform and a fundamental prima facie duty not to perform, the more stringent of the prima facie duties is stronger than the less stringent one.¹⁰ This principle is finite, and it tells us, for each possible action that the agent has both a fundamental prima facie duty to perform and a fundamental prima facie duty not to perform, which duty is stronger. Moreover, there is no reason to think that Ross, or anyone else, would deny it. So, it seems that Ross is not in fact a pluralist(3).

In response, one might try modifying the third tenet of pluralism(3). One might say that the third tenet should be the claim that there is no finite, *non-trivial* principle that can tell us, for each possible action that the agent has both a fundamental prima facie duty to perform and a fundamental prima facie duty not to perform, which duty is stronger. The principle mentioned in the previous paragraph is not a principle of this sort since it is trivial.

However, it's not clear to me that Ross would have accepted even this weaker version of the third tenet of pluralism(3). I suspect that Berker, Gaut, and others assume that Ross accepts something like it because of what he says in the following passage (which we encountered in chapter 1 as well (sec. 1.5)):

¹⁰ Thanks to Fred Feldman for bringing this principle to my attention.

Every act therefore, viewed in some aspects, will be *prima facie* right, and viewed in others, *prima facie* wrong, and right acts can be distinguished from wrong acts only as being those which, of all those possible for the agent in the circumstances, have the greatest balance of *prima facie* rightness, in those respects in which they are *prima facie* right, over their *prima facie* wrongness, in those respects in which they are *prima facie* wrong—*prima facie* rightness and wrongness being understood in the sense previously explained. For the estimation of the comparative stringency of these *prima facie* obligations no general rules can, so far as I can see, be laid down. We can only say that a great deal of stringency belongs to the duties of “perfect obligation”—the duties of keeping our promises, of repairing wrongs we have done, and of returning the equivalent of services we have received. For the rest, [“the decision rests with perception”]. This sense of our particular duty in particular circumstances, preceded and informed by the fullest reflection we can bestow on the act in all its bearings, is highly fallible, but it is the only guide we have to our duty. (1930, 41-42)

One might think that Ross is here suggesting that there is no rule that tells us, for each action that involves a conflict of *prima facie* duties, the relative stringencies of the duties involved. However, I don’t think this passage gives us good grounds for thinking that Ross accepts anything quite so strong. Ross merely says that *so far as he can see*, no general rules regarding the comparative stringencies of conflicting *prima facie* duties can be laid down. He therefore seems to be making primarily an epistemological point. He’s just saying that he doesn’t know of any such rule, not that there is in fact no such rule to be found. Plus, the context of the passage suggests that the types of rules that Ross has in mind are simple ones, ones that would be useful as a “guide” to our duty. For instance, Ross seems to be ruling out principles such as ones that claim that the *prima facie* duty of promise keeping is always more stringent than the *prima facie* duty of benevolence, or that the *prima facie* duty of non-maleficence is always more stringent than the *prima facie* duty of reparation. Ross seems to be suggesting that there are no simple principles such as these, and we therefore have no choice but to use our individual judgment to help us

decide the relative stringencies of the prima facie duties involved in cases of conflict.

However, I don't see anything in the passage that suggests that Ross would deny that it's *in principle* possible—if we had God's knowledge perhaps—to formulate extremely complex (but finite) rules concerning the comparative stringencies of prima facie duties, principles that would, because of their complexity, provide us with little guidance in cases of conflict. I therefore see little motivation for thinking that Ross accepts the third tenet of pluralism(3), contrary to what many philosophers have assumed.

3.5. Pluralism(2)

While pluralism(3) has three distinct tenets, another version of pluralism has only two. Brad Hooker, for instance, says that pluralism is the combination of the following two claims:

1. There is a plurality of moral values or principles.
2. There is no principle that underlies and provides justification for them. (2000, 105; see also 1996)

A few comments on these two tenets. Hooker often speaks as if the first tenet is equivalent to the claim that there is a plurality of “general duties,” which for him seem to be fundamental prima facie duties. So, I'll assume that the first tenet of Hooker's conception of pluralism amounts to the idea that there are multiple fundamental prima facie duties. With regard to the second tenet, Hooker seems to have in mind the idea that there is no explanation for why we have the particular fundamental prima facie duties that we do in fact have (see, e.g., 2000, 106-7). Suppose someone thinks that we have fundamental prima facie duties to keep our promises, to promote the good, and to make

reparation for past wrongs. Why, we might ask, do we have *these* fundamental prima facie duties? Why don't we have other fundamental prima facie duties instead? As I see it, Hooker's second tenet of pluralism claims that no answer to these questions can be given. The fact that we have these fundamental prima facie duties is just a brute fact about the nature of the universe. So, then, what I'll call "pluralism(2)" ("2" for "2 Tenets") is the view that (1) there are multiple fundamental prima facie duties, and (2) there is no explanation for why we have the particular fundamental prima facie duties that we do in fact have.¹¹ Monism(2) is therefore simply the denial of at least one of these claims.

Hooker claims that Ross accepts pluralism(2). In fact, Hooker believes that his own moral theory (a version of rule consequentialism) is superior to Ross's precisely because it rejects pluralism(2) while Ross's accepts it (Hooker 2000, 19-23, 104-7; see also 1996, 2008). Hooker believes that his rule consequentialism gives us an explanation for why we have the particular fundamental prima facie duties that we do in fact have. It therefore denies the second tenet of pluralism(2). But since Ross's theory is committed to pluralism(2) and thus to its second tenet, Hooker claims that his moral theory has a theoretical advantage over Ross's: it is "*more informative and integrated*" (2000, 21) and

¹¹ Other philosophers also understand pluralism in this way. Mark Timmons, for instance, says, "One can be a pluralist about the nature of right action, about the nature of intrinsic value, or about both. If we focus for the time being on the nature of right action, *moral pluralism* involves two main claims: (1) There is a plurality of basic moral rules. (2) There is no underlying moral principle that serves to justify these moral rules" (2002, 190). Similarly, Russ Shafer-Landau claims that pluralism is the view that "there is a plurality of fundamental moral rules" (2012, 215), and he says that a moral rule is "fundamental" when "there are no deeper, more basic moral rules that justify [it]" (2012, 214). See also Stratton-Lake (2002, xii-xiii).

has “more systematic unity” (107). But since the two theories are on a par in every other way, Hooker concludes that his theory is preferable to Ross’s.

But does Ross really accept pluralism(2)? He certainly accepts the first tenet of it. But what about the second? Does Ross really deny that there is an explanation for why we have the particular fundamental *prima facie* duties that we do in fact have? As far as I can tell, there is no passage from *R&G* (or from *FE*) where Ross denies this. If anything, he suggests that he is open to there being such an explanation. Consider what he says immediately after presenting his well-known list of *prima facie* duties:

If the objection be made, that this catalogue of the main types of duty is an unsystematic one resting on no logical principle, it may be replied, first, that it makes no claim to being ultimate. It is a *prima facie* classification of the duties which reflection on our moral convictions seems actually to reveal...If further reflection discovers a perfect logical basis for this or for a better classification, so much the better. (1930, 23)

It’s not entirely clear what Ross means by “a perfect logical basis” and a “logical principle” in this passage. But if he means “an explanation for our fundamental *prima facie* duties,” then he seems to be expressing openness toward the idea that there is an explanation for why we have the fundamental *prima facie* duties on his list.

Of course, this passage is difficult to interpret. Maybe it should be interpreted in some other way. Still, my point remains: there is no good reason to think that Ross accepts the second tenet of pluralism(2). The issue of whether there is an explanation for why we have the particular fundamental *prima facie* duties that we do in fact have is a deeply metaphysical one, and one that, as far as I can tell, Ross ignores. It is thus

probably best to withhold judgment about whether Ross accepts or denies that there is such an explanation, and thus about whether he is a pluralist(2).

3.6. Conclusion

I have now considered four ways of understanding what pluralism is, and I have argued that on each, Ross is not in fact a pluralist, or we have little reason to think that he is. Of course, it may be possible to come up with other conceptions of pluralism according to which Ross *is* a pluralist. One could say, for instance, that pluralism is simply the view that there are multiple fundamental prima facie duties.¹² If that's all pluralism is, then Ross will, of course, be a pluralist. My general point, however, will still hold: Ross is said by many philosophers to be a pluralist; however, if we take a careful look at how these philosophers actually understand pluralism, it's easy to see that Ross is not a pluralist at all. When philosophers say that Ross is a pluralist, they are very often wrong.

¹² This may be how Audi (1996) and Dancy (1983, 1991a, 1991b) understand pluralism.

CHAPTER 4

ROSS'S OBJECTIONS TO ROSS

4.1. Introduction

In chapter 1, I argued that in *R&G*, Ross accepts the following view of moral obligation:

RTO_{RG}: An action *A* is morally obligatory for *S(A)* to perform if and only if (and because) *S(A)* can perform *A*, and *A*'s total prima facie rightness minus its total prima facie wrongness is greater than that of any other action *S(A)* can perform instead.

However, in Ross's later book, *FE*, he changes his view significantly. In this chapter, I will begin by discussing the details of Ross's new view. After that, I will examine the arguments Ross gives for adopting it instead of the view he held in *R&G*. I will argue, however, that none of these arguments is persuasive.

4.2. Ross's Theory of Moral Obligation in *FE*

Ross devotes chapter 3 of *FE* to non-naturalistic definitions of the term "right"—in other words, he discusses various attempts to provide a definition of "right" that makes use of other normative terms. At the end of the chapter (1939, 51-56), he discusses the view of C. D. Broad. According to Broad, "right" means, in general, "appropriate, fitting, or suitable to the situation," and "*morally* right" means, in particular, "*morally* appropriate, fitting, or suitable to the situation." Of course, one will wonder what Broad means by "morally appropriate, fitting, or suitable" in this context. However, according to Broad, these terms cannot be further defined: the concept they express is *sui generis*.

Ross is sympathetic with Broad's account, but Ross thinks it needs to be slightly modified (1939, 52-53). He has us imagine a person with two alternatives, each of which is morally suitable to the situation, but one is more suitable than the other. If "morally right" meant "morally suitable to the situation," then each alternative would be right. But that's not the case: the more suitable option is the only right one in the scenario. So, Ross suggests, instead of saying that "morally right" means "morally suitable to the situation," it would be more accurate to say that "morally right" means "*most* morally suitable to the situation."

Later on, in chapter 7 of *FE*, Ross reaffirms his general acceptance of Broad's view of the definition of "right." He says, "Let me start by supposing that we accept the general account of rightness that has been offered, viz. that '[morally] right' means 'suitable, in a unique and indefinable way which we may express by the phrase "morally suitable," to the situation in which an agent finds himself"' (1939, 146). But Ross then adds something new to the account. He claims that a person's situation contains two elements: the "objective" element and the "subjective" element (146). Actions that are morally suitable to the objective element of the agent's situation are morally right in the objective sense of the term, and actions that are morally suitable to the subjective element of the agent's situation are morally right in the subjective sense of the term.

According to Ross, the objective element of a person's situation consists of the non-mental facts of the situation. As he puts it,

The objective element consists of facts about the various persons and things involved in the situation, in virtue of which a certain act would in fact be the best possible fulfillment of the various *prima facie* obligations resting on the agent. Suppose, for instance, that the situation is one in which none of the special obligations such as that

of keeping a promise or of making reparation for an injury rests upon the agent, but only the responsibility for bringing as much good as possible into being. Then the act which would in fact produce the maximum of good will be that which best fits the objective element of the situation, and will be in this respect the right act. (1939, 146)¹

Here Ross seems to be suggesting something like the following theory of the objective sense of “morally right”:

R3: An action A is morally right (in the objective sense) for S(A) to perform if and only if (and because) S(A) can perform A,² and A’s total prima facie rightness minus its total prima facie wrongness is greater than that of any other action S(A) can perform instead.

In contrast to the objective element, the subjective element of an agent’s situation consists of the mental facts of the situation (in particular, the agent’s beliefs). As Ross puts it,

The subjective element consists of the agent’s thoughts about the situation. These are as much parts of the total situation as are the objective facts. And the act which is morally suitable to them, i.e. the act which the agent, in view of his opinion about the situation, thinks will be the maximum fulfillment of [prima facie] obligation, will be in that respect right. (146)

¹ See also (1939, 165): “If by a right act we mean an objectively right act, i.e., the act which out of all those open to a particular agent in particular circumstances will in fact produce the maximum fulfillment of the claims that exist against him, we must maintain the complete non-dependence of moral goodness and rightness upon one another.”

² Although this clause is not explicitly mentioned in the quoted passage, I think Ross would accept it. He claims that while there is a “wide” sense of “right” according to which it is possible for an emotion to be right even though the person has no control over it, this is not the case when we consider actions instead of emotions. Ross says, “it must be granted that when we use ‘right’ of acts, as opposed to emotions, we usually think of them as being in the agent’s power to do or to forbear from doing” (1939, 55).

Here Ross seems to be suggesting the following theory of the subjective sense of “morally right”:

R4: An action A is morally right (in the subjective sense) for S(A) to perform if and only if (and because) S(A) can perform A,³ and S(A) believes that A’s total prima facie rightness minus its total prima facie wrongness is greater than that of any other action she can perform instead.

Next, Ross stresses that it’s important to distinguish between the objective sense and the subjective sense of “morally right”: failure to do so has created much confusion and has caused many who are engaged in moral discourse to talk past each other (1939, 146-47).⁴ However, Ross acknowledges that if we posit the existence of multiple senses of “right,” one will be bound to wonder which of them is superior to the others. He says,

The recognition of the difference between the two [senses of “right”] is therefore in itself important as tending to reconcile what might otherwise seem irreconcilable difference of opinion. But the question remains, which of the characteristics—objective or subjective rightness—is ethically the more important, which of the two acts is that which we ought to do. (147)

Ross claims that he used to think that the objective sense of “right” is more important than the subjective sense. In other words, he used to think that we ought to do what is right in the objective sense. Ross calls this view the “objective view.” That Ross used to accept the objective view is clear, given his prior commitment to RTO_{RG} . But Ross says that he has now shifted his position. Now, he says, he has “come to hold the opposite

³ Again, although Ross doesn’t explicitly mention this condition in the passage cited, I think he’d accept it (see note 2).

⁴ In fact, Ross seems to think that there are more than two senses of “morally right”: he claims that an action that is morally suitable to the agent’s motives is morally right in a third sense of the term (1939, 159).

opinion, that it is the subjectively right act that is obligatory” (148). In other words, Ross says that he has come to accept the “subjective view” rather than the objective view—he has come to hold that it is the subjectively right action, not the objectively right action, that is obligatory.⁵ Given this, and given his commitment to R4, it may seem that whereas Ross used to accept RTO_{RG} , he now accepts the following:

R5: An action A is morally obligatory for S(A) to perform if and only if (and because) S(A) can perform A, and S(A) believes that A’s total prima facie rightness minus its total prima facie wrongness is greater than that of any other action S(A) can perform instead.

(R4 is identical to R5 except that “subjectively right” in R4 has been replaced by “obligatory” in R5.)

Although R5 seems to be the theory of obligation that Ross accepts at the beginning of chapter 7 of *FE*, he makes an important modification to it midway through the chapter. According to R5, *actions* are the sorts of things that can be obligatory, but Ross claims to

⁵ Ross’s position—that “right” has an objective and a subjective sense and that subjectively right actions, not objectively right ones, are obligatory—is puzzling for the following reason. If Ross is using the term “right” in this context to mean “permissible,” then it’s strange that he’d identify any sense of “right” with “obligatory”—after all, permissibility and obligatoriness are different concepts; an act can be permissible without being obligatory, as Ross points out (1930, 3-4; 1939, 43-45). On the other hand, if Ross is using the term “right” in this context to mean “obligatory,” as he typically does in *R&G*, then his claim that objectively right actions are not obligatory seems incoherent. Perhaps, then, when Ross says that there are two senses of “right,” he means that there are two senses of “obligatory,” and when he asks whether we “ought” to perform the objectively right act or the subjectively right act, the ought he’s appealing to here is a different sort of ought altogether, something like an all-things-considered ought, or a just-plain ought (for more on this type of ought, see McLeod [2001]). If that’s what Ross intends, then, since he also takes himself to have changed his position, this just-plain ought must be the same ought that he was writing about in *R&G*, and, in particular, it must be the same ought at issue in RTO_{RG} .

have been convinced by Prichard that this isn't, strictly speaking, correct: we are only ever obligated to *set ourselves* to do things. As Ross says,

Professor Prichard now turns to consider a difficulty which is common to both views [i.e., the view that we are obligated to do what is right in the subjective sense, and the view that we are obligated to do what is right in the objective sense], and which if well founded will lead us to modify both. We have hitherto assumed that an obligation is an obligation to do some *action*, i.e. to produce some change in something. But we must ask whether this is true. (1939, 153)

The most important point, I think, which emerges from Professor Prichard's discussion is that the only thing to which a man can be morally obliged is what I will call a self-exertion, a setting oneself to effect this or that change or set of changes. He cannot be obliged to perform an "act," in the ordinary sense. (160)⁶

Why does Ross think that self-exertions, not actions, are the objects of obligation?

His argument is murky,⁷ but it seems to be something like the following. Recall that according to Ross, an action consists of a self-exertion having an effect (see sec. 1.4). Thus, the action of my waving my hand consists of a self-exertion of mine having the effect of my hand's waving. However, Ross argues that we never have control over the effects of our self-exertions. For instance, my setting myself to wave my hand often has the effect of my hand's waving, but not always. I might set myself to wave my hand without its actually waving, if, for instance, I don't realize that I suffer from a certain sort of paralysis. Since we never have control over the effects of our self-exertions, and since our actions consist, in part, of the effects of our self-exertions, it follows that we never

⁶ Ross (1939, 108, 160-61) makes it clear that this applies to our prima facie obligations as well: we are never prima facie obligated to act; we are only ever prima facie obligated to set ourselves to act.

⁷ Ross's argument appears in (1939, 153-54, 160-61).

have control over our actions. However, Ross thinks that we can be obligated to do only what is in our control. Because of this, we cannot be obligated to perform actions. But self-exertions are unlike actions in this respect. Self-exertions do not consist of effects that are outside of our control. Because of this, we have control over our self-exertions, and we should therefore regard them, not actions, as the objects of obligation.

As Michael Zimmerman (2008, 11-13) has pointed out, this argument is problematic. First, Zimmerman claims, we *do* have a certain type of control over the effects of our self-exertions. We have *partial* control over them, as Zimmerman calls it. Partial control contrasts with complete control. As Zimmerman describes these notions, “One has complete control over something only if its occurrence is not contingent on anything else that is beyond one’s control; otherwise, any control one has over it is merely partial” (2008, 12). Since the effects of our self-exertions are contingent on our self-exertions, which we do have control over (at least according to Ross), it follows that we have partial control over these effects. And given how Ross conceives of actions, it also follows that we have partial control over our actions. Thus, Ross’s claim that we don’t have control over our actions is false if “control” means “partial control.”

Perhaps, then, when Ross uses “control” in the above argument, he means “complete control.” But then, unless he is equivocating, the premise that we can be obligated to do only what is in our control is problematic. As Zimmerman says,

It is clear that no one ever has complete control over anything, including any and all self-exertions. (If you doubt this, consider the simple fact that whatever control you enjoy over anything depends on your having been born—something that we may hope was in someone’s control, but not yours. Succeeding in exerting oneself, just like succeeding in doing that to which one exerts oneself, requires the co-operation of all sorts of factors, both past and present, over which one lacks control). If obligation

required complete control, nothing would be obligatory. Since (I assume) some things are obligatory, we may conclude that obligation does not require complete control. (2008, 13)

If one doubts Zimmerman's reasons for thinking that we lack complete control over our self-exertions, perhaps the following will bolster the case. As we've seen, Ross thinks that we lack (complete) control over the effects of our self-exertions because we might suffer from unknown paralyses: I might set myself to wave my hand, but because of an unknown paralysis, my self-exertion might not cause my hand to wave. However, it seems that a similar thing could happen for self-exertions. Suppose, for instance, that the neural pathway in my brain that must be activated for me to set myself to wave my hand has been damaged, but I don't realize this. In that case, I might think that I can set myself to wave my hand, but because I suffer from an unknown paralysis of self-exertion, I am unable to actually do so. It seems, then, that my setting myself to do something is contingent on no paralysis like this from occurring, and I therefore lack complete control over even my self-exertions.

These arguments aside, the Ross of *FE* clearly holds that we are obligated to perform self-exertions, not actions, and so he would presumably urge us to modify R5 to something like this:

R6: A self-exertion E is morally obligatory for S(E)⁸ to perform if and only if (and because) S(E) can perform E, and S(E) believes that E's total prima facie rightness minus its total prima facie wrongness is greater than that of any other self-exertion she can perform instead.

⁸ "S(E)" refers to the agent of self-exertion E.

Is R6 Ross's considered view in *FE*? Perhaps. However, toward the end of chapter 7, Ross seems to state his view in a somewhat different way. After claiming that self-exertions are the only things that one can be obligated to perform, he says that there are "several different self-exertions which might have some claim to be considered right, or what the agent ought to do." First, there is the "self-exertion which is morally most suitable to the objective circumstances, in the sense of 'circumstances other than the agent's own state of knowledge or opinion.'" Second, there is the "self-exertion which is morally most suitable to the agent's state of mind about the circumstances, in which there may be included ignorance and false opinion as well as knowledge and true opinion." Finally, there is the "self-exertion which he *thinks* to be morally most suitable in the circumstances as he takes them to be" (1939, 161). Ross makes it clear that his position is the third one (162). Given this, it may seem that Ross's view in *FE* is the following:

R7: A self-exertion E is morally obligatory for S(E) to perform if and only if (and because) S(E) can perform E, and S(E) believes that E is the most morally suitable thing he can do in the circumstances he thinks he's in.

R7 is importantly different from R6. According to R6, a self-exertion is obligatory only if the agent believes its total prima facie rightness over wrongness is greater than that of its alternatives. But according to R7, a self-exertion can be obligatory even if the agent has no view about its total prima facie rightness over wrongness—the agent, rather, must have a belief about its "suitability" to the situation.

So which view is Ross's, R6 or R7? It's not clear to me. I think the text is indeterminate on the matter. However, no matter which theory of obligation Ross accepts in *FE*, we are now in a position to appreciate why he accepts it. In the remainder

of this chapter, I will examine Ross's arguments⁹ for rejecting objective views such as $R_{TO_{RG}}$ and accepting a subjective view like R6 or R7 instead.

4.3. Argument #1

Ross says, "much of our ordinary thought is in conflict with the objective view. (A) One instance will suffice to show this. Suppose one is driving a car from a side-road into a main road; the question arises, ought one to slow down before entering the main road." Ross then quotes Prichard's contention that "If the objective view be right, (1) there will be a duty to slow down only if in fact there is traffic." Ross concludes that "Once this is realized it becomes clear that most of our ordinary thought involves the subjective view" (1939, 151-52).

In order to better understand the argument Ross is making here, it will be useful to consider a somewhat more detailed version of the example he appeals to:

Lane Merging

I am about to merge onto a normally very busy and dangerous highway. Having merged onto this highway many times before, I know about its dangers. Suppose I have two options: I can slow down and look for traffic before merging, or I can speed up and merge without looking. I don't know it, but if I slow down and look for traffic, this will cause me to be late for an important meeting, but if I speed up and disregard traffic (which happens to be very light today), I will arrive safely at the meeting on time—no one will be harmed.

⁹ As Ross makes clear, most of these arguments are either taken directly from, or are heavily inspired by, Prichard ([1932] 2002).

Ross's point seems to be that if an objective theory is true—that is, if a theory that holds that the moral obligatoriness of an action is determined solely by, as Ross puts it, “the facts about the various persons and things involved in the situation” is true—then I ought to merge recklessly. But this implication is counterintuitive. Thus, no objective theory is true, and a subjective theory must be true instead—that is, a theory that holds that the moral obligatoriness of an action is determined by, as Ross puts it, “the agent's thoughts about the situation” must be true.

Stated more formally, Ross's argument looks like this:

- 1) If an objective theory is true, then I am obligated to merge recklessly.
- 2) But I am not obligated to merge recklessly.
- 3) So, no objective theory is true.
- 4) If (3), then a subjective theory is true.
- 5) So, a subjective theory is true.

There are several problems with this argument. First, premise (1) is very strong: according to it, *every* objective theory implies that I am obligated to merge recklessly. But that just seems false. Consider, for instance, the following theory: an action is morally obligatory iff it involves slowing down and looking for traffic. This theory appears to be an objective theory—according to it, the moral obligatoriness of an action is determined solely by the facts of the agent's situation. However, the theory obviously does not imply that I should merge recklessly in Lane Merging. Of course, this theory is a silly one; no one would ever adopt it. But even plausible objective theories may not imply that I ought to merge recklessly. Suppose, for instance, that in Lane Merging, I have made a promise to my wife to drive safely. Perhaps I've even specifically promised her that I will look for traffic as I merge onto the highway in question. In that case,

Ross's objective theory (RTO_{RG}) may imply that I ought to merge carefully (it will imply this if my prima facie duty to keep my promise outweighs my other prima facie duties). So, even plausible objective theories may not have the supposedly unintuitive implication that Ross suggests—whether they will have it depends on how the details of the case are filled in.

Ross's argument has other problems. In particular, premise (2) is controversial. Ross supports this premise by appealing to “ordinary thought”; he suggests that ordinary thought (i.e., commonsense morality) dictates that I should merge carefully, not recklessly. However, a proponent of an objective view—an “objectivist”—can respond to this in the following way. She can say that I *am* obligated to merge recklessly in the case. However, it might *seem* like I'm not because I wouldn't be praiseworthy for doing so, and it's easy to confuse lack of praiseworthiness with lack of obligation. Indeed, ordinary moral thought likely makes no distinction between obligation and praiseworthiness. However, we must be careful to distinguish between these concepts. Once we do, it will no longer seem so obvious that premise (2) is true.

Of course, this maneuver is nothing new. Utilitarians have long distinguished between praiseworthiness/blameworthiness and obligatoriness/wrongness in order to respond to arguments like the present one (see, e.g., Moore [1912] 2005, 100-1). In fact, in several places in *R&G*, Ross distinguishes between obligatoriness and what he calls “moral goodness” (which he, in turn, seems to identify with moral praiseworthiness), and he uses the distinction to combat various objections (1930, 32, 45).¹⁰ Moreover, Ross

¹⁰ I will discuss Ross's view of moral goodness in detail in chapter 8.

continues to differentiate these concepts in *FE*, and he admits that moral goodness and moral rightness can come apart, no matter whether an objective or a subjective theory is true (1939, 163). For this reason, it's strange that Ross does not consider a response to his lane-merging argument that appeals to such a distinction. He seems to no longer find this type of response convincing (1939, 147-48), but it's unclear why. Because of this, I think we must conclude that Ross has failed to adequately support premise (2) of the argument: there is a plausible way of objecting to the premise that Ross leaves unchallenged.

There is a final problem with Ross's argument that deserves mentioning. If sound, his argument shows that subjectivism is true. However, it doesn't show that any particular subjective theory is true. It doesn't, in particular, show that *Ross's* subjective theory (understood as either R6 or R7) is true. In fact, one might reasonably wonder whether either R6 or R7 yields the implications that Ross wants it to yield. R6, after all, doesn't imply that I ought to merge carefully, for R6 is only a theory about when a *self-exertion* is obligatory. At most, it implies that I am obligated to set myself to merge carefully. And it will imply that only if I believe that the total prima facie rightness minus the total prima facie wrongness of setting myself to merge carefully is greater than that of any other thing I can set myself to do instead. But I might very well lack this belief at the time of my action. Maybe I'm a utilitarian, not a proponent of R6, and I thus spend little or no time thinking about which of my self-exertions maximize the balance of total prima facie rightness over wrongness. Or maybe I do subscribe to R6 and do sometimes form beliefs about the prima facie rightness and wrongness of the things I can set myself to do; however, as I am merging onto the highway, I am thinking about other

things, and no thought about the prima facie rightness and wrongness of my self-exertions crosses my mind. Similar things can be said, I think, about R7. For these reasons, it's not clear to me that either R6 or R7 will in fact yield Ross's intended implications about Lane Merging. So, if the case causes problems for objective theories, then it may very well cause problems for Ross's subjective theory too.

4.4. Argument #2

According to Ross, "There is another mode of argument by which we may, I think, satisfy ourselves of the truth of the subjective view. It might be agreed, I believe, that the act which a man in any situation ought to do is that which it would be *reasonable* for him to do if he wanted to do his duty in that situation" (1939, 156). However, Ross thinks that if an objective theory of obligation is true, then it will be possible for unreasonable actions to be obligatory. To illustrate this, he gives an example (156). Suppose a friend has lent me a book, and I promise to mail it back to him after I'm finished with it. Suppose I finish the book and have two options: I can "dispatch" the book carefully (e.g., I can put it in a clearly labeled box and mail it with the correct postage attached), or I can dispatch the book carelessly (e.g., I can put it in an unaddressed box and mail it without postage). Unbeknownst to me, if I dispatch the book carefully, it will never reach its destination (perhaps the mail truck will get into an accident and its cargo will be destroyed). On the other hand, if I dispatch the book carelessly, it will, miraculously, reach my friend (the reader is free to imagine how exactly this might happen). Ross suggests that the latter option "would be objectively right" since it "would succeed in returning to a friend a book [I] had promised to return," even though this consequence is

one that “no human foresight could foresee” (156). However, Ross claims that “if no human foresight can foresee these facts, no one would say that it was reasonable for a man who wanted to do his duty by his friend to dispatch the book carelessly, since the successful result of this neither is nor could be foreseen by the sender” (157).

The argument that Ross seems to be making here can be stated like this:

- 1) If an objective theory is true, then I ought to dispatch the book carelessly.
- 2) But it would be unreasonable for me to dispatch the book carelessly.
- 3) So, if an objective theory is true, then it can happen that a person is obligated to do something that’s unreasonable for her to do.
- 4) But no one can be obligated to do something that’s unreasonable for her to do.
- 5) So, no objective theory is true.
- 6) If (5), then a subjective theory is true.
- 7) So, a subjective theory is true.

This argument shares many of the problems of Ross’s lane-merging argument.

Premise (1) is unwarrantedly strong. Some objective theories won’t imply that I ought to dispatch the book carelessly (consider, for example, the view that an action is obligatory iff it is the mailing of a package in a clearly labeled box with the correct postage affixed). In addition, depending on how the details of the case are filled in, other, more plausible, objective theories may also fail to have this implication. Moreover, I suspect that on some versions of the case, even some subjective theories will not imply that I ought to dispatch the book carefully. Perhaps when I mail the book, I don’t happen to think about the *prima facie* rightness or the “moral suitability” of my available actions (or self-exertions). If that’s the case, then Ross’s subjective theory, under either of its guises, will not imply that I ought to mail the book carefully.

However, Ross's argument has a more serious problem: it's unclear how he's using the term "reasonable" in the argument.¹¹ When Ross says that an action is reasonable, perhaps he just means that there is a reason for the agent to perform it. But if that's the case, then an objectivist can plausibly deny premise (2). On this interpretation, premise (2) amounts to the claim that there is no reason for me to dispatch the book carelessly. But an objectivist can plausibly maintain that there *is* a reason for me to dispatch the book carelessly—after all, doing so fulfills a promise I made to my friend.

Perhaps, though, when Ross says that an action is reasonable, he doesn't merely mean that there *is* a reason for the agent to perform it; rather, he means that the action is one that the agent *has* a reason to perform. The difference between there *being* a reason for a person to do something and the person's *having* a reason to do it is this. To say that there is a reason for a person to perform an action just is to say that there is some consideration that counts in favor of the action's performance. On the other hand, to say that a person has a reason to perform an action is to say (roughly) that there is a reason for the person to perform the action *and* the person is in some way aware of it.¹² To illustrate, suppose that one of my alternatives will have some fabulous consequence in the distant future. If that's the case, then there surely *is* a reason for me to perform the action. Suppose, in addition, that I know that the action will have this future consequence. In that case, then there not only is a reason for me to perform the action, but it's also one that I have. On the other hand, suppose that I do not know that the action

¹¹ McConnell (1998, 85-86) and Zimmerman (2008, 9) make a similar observation.

¹² For why this is only roughly correct, see Schroeder (2008).

will have this future consequence. Then, although there is a reason for me to perform the action, I do not *have* it.

On this interpretation of “reasonable,” premise (2) amounts to the claim that I have no reason to dispatch the book carelessly. This is plausible. Although there is a reason for me to dispatch the book carelessly since doing so will fulfill my promise, it’s not a reason that I am aware of. However, on this interpretation, premise (4) is question begging, or at least close to it. On this interpretation, (4) amounts to the claim that no one can be obligated to do something that she has no reason to do. However, to have a reason to perform an action in the sense at issue here involves being *aware* of a consideration that counts in favor of the action. But being aware of such a consideration involves having a *belief* about that consideration. Thus, according to premise (4), beliefs are relevant to the determination of a person’s obligations. But whether such a claim is true is precisely what is at issue between objectivists and subjectivists. To include such a claim in an argument against objective theories is to beg the question against the objectivist.

However, perhaps Ross means something different when he uses the term “reasonable” in his argument. Maybe when he says that an action is “reasonable,” he simply means that it “makes sense” for the agent to do it. To illustrate, suppose I’ve had a long day at work, and when I get home, what I’d most like to do is relax. In addition, suppose I believe that listening to music will allow me to relax, but reading a book will not. In this case, it would, *ceteris paribus*, “make sense” for me to listen to music, and it would “make no sense” for me to read a book instead. The former action is therefore reasonable on the current sense of the term, but the latter action is not.

On this interpretation, premise (2) again looks beyond reproach. After all, in Ross's book-dispatching case, I presumably want to return the book to my friend in a timely fashion, and I believe that sending the book to him in a clearly labeled box with the correct postage is a good way of achieving that goal. On the other hand, I presumably believe that sending the book in an unlabeled box with no postage is a terrible way of achieving my end. Given this, if I were to send the book in the former way, my action would make perfect sense for me to do. It would thus be reasonable in the sense at issue. On the other hand, if I were to send the book in the latter way, my action would be bizarre. It wouldn't make any sense for me to do. It would therefore be unreasonable in the sense at issue.

But if this is what Ross means by "reasonable," then while premise (2) is plausible, premise (4) again looks question begging. On the current interpretation of "reasonable," premise (4) amounts to the claim that no one can be obligated to do something that doesn't make sense for her to do. But as the phrase is being used in this context, whether an action "makes sense" clearly depends on the agent's subjective state—in particular, on her beliefs and her ends. Thus, to presume that obligatory actions must "make sense" for the agent to do is to beg the question under dispute.

A final suggestion. I think that sometimes when people say that an action is reasonable, they just mean that the action is permissible. If that is what Ross means by "reasonable," then premise (4) is surely true: no one can be obligated to do something that's impermissible for her to do. But now premise (2) is highly contentious. On this interpretation of "reasonable," (2) amounts to the claim that it would be impermissible for me to dispatch the book carelessly. Of course, an objectivist will not agree to that! Ross

might retort that it is counterintuitive to maintain that I am permitted to send the book carelessly. But now the argument collapses into the argument from the previous section. An objectivist can respond that I *am* permitted to send the book carelessly, but I wouldn't be praiseworthy for doing so. As we've seen, Ross has done nothing to invalidate this argumentative strategy.

Perhaps there is some other way of interpreting Ross's use of "reasonable" according to which each of the premises of his book-dispatching argument is plausible (and none beg the question). However, I know of no interpretation of the term according to which this is the case. For that reason, I find the argument to be unpersuasive.

4.5. Argument #3

Consider what Ross says here:

Yet this view [the objective view] has awkward consequences. (1)¹³ In order to know that some moral rule is applicable to me here and now, I must know (*a*) that the situation contains a thing of the kind *A* capable of having a state of the kind *x* effected in it, and (*b*) that it is such that some act that I can do would cause this *A* to assume a state of the kind *x*. Now (*a*) is not always fulfilled. I may not know whether my parents are in difficulties, or whether a man I meet is ill. And (*b*) is *never* fulfilled; I never *know* nor can come to *know* that some state which I can bring about will produce an effect of the kind *x*, though I may have reason to *think* it. Thus if duty be such as the objective theory conceives it to be, I can never know that I have any particular duty, or even that any one has ever had or will ever have a duty. (1939, 149-50; underlining added)

This is a difficult passage. However, the general structure of Ross's argument is clear:

¹³ In the next section, I will discuss the other "awkward consequences" that Ross believes the objective view has.

- 1) If an objective theory is, then I never know what I am obligated to do.
- 2) But I do know, at least sometimes, what I am obligated to do.
- 3) So, no objective theory is true.

(Note that in contrast to Ross's previous arguments, this argument is directed solely at disputing objective views. However, Ross would presumably be happy to add the following lines to the argument: (4) If no objective theory is true, then a subjective theory is true; (5) So, a subjective theory is true.)

While the basic outline of Ross's argument may be easy to ascertain, it is much more difficult to reconstruct his rationales for its premises. Ross does not explicitly mention premise (2) and gives no rationale for it. Perhaps he'd say that it is a dictum of commonsense morality. Ross does give a rationale for premise (1), but his reasoning is difficult to follow in the cited passage. To better understand what Ross has in mind, it's helpful to consider what he says in the passage that immediately follows:

It is worth while to note in passing just what is proved and what is not proved by this argument. In constructing or in following a geometrical proof, we never know that we have before us a triangle, for instance; but we treat the figure before us as if it were a triangle, and we come to *know* that if it were, it would have certain properties. Similarly, we never know that an act we could do would produce a certain effect, but we may think that it would, and may *know* that if it would, it would be our duty to do it; and we might proceed from this to generalized moral rules, which would be hypothetical in character; e.g. "if you can ever produce a true opinion in the mind of some one else as to what you think, you ought to do so." Thus the objective view is not fatal to the possibility of knowing moral rules. But it is fatal to the possibility of recognizing particular duties incumbent on us here and now, since we can never know, for instance, that we *can* produce a true opinion as to our thought in any one else's mind. (1939, 150; underlining added)

Ross's thought seems to be this. All objective theories claim that obligatory actions are ones that have some kind of non-mental effect or feature. This feature, according to

RTO_{RG} , is the maximization of prima facie rightness over wrongness. According to traditional act utilitarianism—a paradigmatic objective theory—the feature is the maximization of utility. According to divine command theory—another good example of an objective theory—it’s the conformity to God’s commands. And so forth. But if any of these theories is true, then in order to know that an action of ours is obligatory, we have to know that the action has or will have the effect or feature in question. Thus if RTO_{RG} is true, then we have to know that an action of ours maximizes the balance of prima facie duty in order to know that it is obligatory. If utilitarianism is true, then we have to know that an action of ours maximizes utility in order to know that it is obligatory. Etc. But we can never know these things. It therefore follows that if an objective theory is true, we never know what it is that we ought to do.

However, this line of reasoning is problematic. First, one can challenge Ross’s suggestion that we can never know that an action of ours has the effect or feature specified by an objective theory (e.g., the feature of maximizing the balance of prima facie rightness over wrongness, or the feature of maximizing utility). It might be that we can never be *absolutely certain* that an action of ours has such a feature. But one can plausibly deny that absolute certainty is required for knowledge. And if absolute certainty is not necessary for knowledge, then perhaps we can, at least on occasion, be *certain enough* that an action of ours has the effect or feature in question to *know* that it does. Suppose, for instance, that we find ourselves in a situation like Peter Singer’s shallow pond case (1972). There is a small child drowning in a shallow pond, very near the shore. It would be easy for us to save the child. We merely have to wade into the pond and pull the child to safety. If we do nothing, however, the child will perish. One

can plausibly maintain that in a case such as this, we might very well be sure enough that saving the child maximizes utility (prima facie duty, etc.) to know that it does. Of course, even if this is correct, it's unclear how often we find ourselves in situations relevantly like this. But the general point remains: Ross's claim that we can never know that our actions have the effects or features specified by objective theories can be fairly questioned.

Second, one can challenge Ross's suggestion that if an objective theory is true, then in order to know that an action is obligatory, one must know that it has the feature or effect specified by the theory. (Ross seems to assume that if RTO_{RG} is true, then in order to know that an action is obligatory, we must know that it maximizes the balance of prima facie duty. On the other hand, if utilitarianism is true, then in order to know that an action is obligatory, we must know that it maximizes utility.) Ross is presumably correct that *one* way we might come to know that an action of ours is obligatory, if an objective theory is true, is by deducing that it is obligatory from the theory. But this is surely not the *only* way that we might come to know this. Suppose, for instance, that a moral expert tells us that an action of ours is obligatory. Suppose we greatly trust this expert, and she has a proven track-record of giving accurate moral advice. If we come to believe that our action is obligatory on the basis of this moral expert's pronouncement, then I think our belief might very well be an instance of knowledge. Or suppose that we have a strong intuition that an action of ours is obligatory. If intuitions can provide justification for beliefs, then if we come to believe that an action of ours is obligatory on the basis of our moral intuition, our belief might again be an instance of knowledge. Finally, perhaps we have discovered a set of easily implementable moral principles (or rules of thumb) to

supplement our objective moral theory (utilitarians have long attempted to provide such principles). If we come to believe that an action of ours is obligatory on the basis of these principles, then once again, our belief might be a case of knowledge.

For the two reasons just mentioned, Ross's rationale for the first premise of his knowledge argument is unconvincing. Moreover, it's difficult to see what a convincing justification for the premise would look like. This premise, then, seems to me to be the argument's undoing. However, before moving on, I'd like to make one further observation about the argument. If, contrary to what I've said, the argument successfully refutes all objective theories, then it (or, at least, a very similar argument) likely refutes many subjective theories as well. This is because many subjective theories do not provide easier access to moral knowledge than objective theories do. Consider, for instance, expected utility utilitarianism. This theory claims that an action is obligatory iff it maximizes expected utility, where the expected utility of an alternative is the sum, for each of its possible outcomes, of the outcome's actual utility times the agent's subjective probability (i.e., the agent's credence, or degree of belief) that the outcome will obtain if the alternative is performed.¹⁴ However, it is no easier to use this theory to determine one's obligations than it is to use traditional act utilitarianism to determine them.¹⁵ After all, in order to use expected utility utilitarianism to determine whether an alternative of

¹⁴ See sec. 2.5 for a more precise formulation of expected utility utilitarianism. Note here that I am assuming that the sort of probability involved in expected utility utilitarianism is *subjective* probability. This is how the view is standardly formulated, though it could be formulated in terms of other types of probability, such as epistemic probability or objective probability (for more on these types of probability, see sec. 7.2).

¹⁵ Feldman (2006) forcefully illustrates this point.

ours is obligatory, we need to know not only the actual utility of each of the alternative's outcomes, but also our subjective probability that each of these outcomes will obtain. And even if we could determine these things, we'd still have a multitude of mathematical calculations to perform. Thus, if the truth of traditional act utilitarianism implies that we never know what we ought to do, then it seems that the truth of expected utility utilitarianism does too.

In response to this, Ross might argue that there is no great problem here, for *his* subjective theory *does* provide an easier path to moral knowledge than traditional act utilitarianism does. After all, if R6 is true (and we know it), then we can come to know whether a self-exertion of ours is obligatory simply by determining what we believe: we simply have to determine whether we believe the self-exertion has a greater balance of prima facie rightness over wrongness than that of any other self-exertion we can perform instead. Likewise for R7: if it's true and we know it, then we can come to know whether an action of ours is obligatory by determining whether we believe it's the morally most suitable one in the situation we think we are in. And while we may not always be able to determine what we believe, we typically can. So, even if some subjective theories give us no greater access to moral knowledge than objective theories do, Ross can argue that this is not the case for his subjective theory.

However, one might wonder at this point whether Ross's theory makes moral knowledge a bit *too* easy. After all, it intuitively seems that we cannot resolve moral quandaries simply by pondering what it is that we believe. This is Zimmerman's complaint about subjective theories like Ross's. He says that this type of view "makes a mockery of the conscientious person's inquiry into what he ought to do, implying that

such an inquiry can be successfully accomplished simply by way of introspection” (2008, 14). I’ll return to this line of thought in sec. 7.2.

4.6. Some Final Arguments

The last of Ross’s reasons for abandoning objective theories like RTO_{RG} appear in the following passage:

Further [awkward] consequences of the objective view are (2)¹⁶ that we can never do a duty because it is a duty, since this must mean “because we know it to be a duty,” (3) that some past act of mine may have been morally obligatory though I believed it was one I ought not to do, and (4) that I may do some act which is obligatory, though I do not even suspect that it will have the effect which renders it a duty. These difficulties all arise from supposing that an act is made my duty by the objective facts of the situation. The only alternative is to suppose that it is made a duty by the subjective facts of the situation, viz. by my state of knowledge or opinion about the facts of the case. (1939, 150)

Here Ross makes several arguments against objective theories (and thus in favor of subjective theories). Here’s one:

- 1a) If an objective theory is true, then a person can never do what she is obligated to do because she is obligated to do it.
- 2a) But the consequent of (1a) is false.
- 3) So, no objective theory is true.
- 4) If no objective theory is true, then a subjective theory is true.
- 5) So, a subjective theory is true.

The other arguments Ross makes in the above passage look just like this one but with a different first (and a slightly different second) premise:

¹⁶ This list is a continuation of the one Ross began on the previous page (see the passage quoted at the beginning of sec. 4.5).

- 1b) If an objective theory is true, then it can be the case that a person was obligated to do something even though she believed she shouldn't do it.
- 2b) But the consequent of (1b) is false.

...

- 1c) If an objective theory is true, then it can be the case that a person is obligated to do something even though she does not suspect that it has the effect or feature that makes it obligatory.
- 2c) But the consequent of (1c) is false.

...

Each of these arguments has many of the problems that Ross's other arguments have, so I can deal with them quickly. Consider premise (1a). From the passage above, it's clear that Ross thinks this premise is true because he thinks that in order to do something because we are obligated to do it, we must know that we are obligated to do it; however, the truth of an objective theory implies that we never know what we are obligated to do. But as I suggested in the previous section, it's not clear that the truth of an objective theory would indeed rob us of all moral knowledge. Premises (1b) and (1c), on the other hand, are more defensible. However, even they are probably too strong. Instead of saying that *every* possible objective theory has the implications specified by (1b) and (1c), it would be more tenable to say only that *most plausible* objective theories do.

Furthermore, premises (2a), (2b), and (2c) are contentious. Consider premise (2b). Why does Ross think that a person cannot be obligated to do what she believes she shouldn't do? It's unclear. However, I suspect that Ross is thinking again of cases like the lane-merging case and the book-dispatching case. Suppose that in these cases, I believe, respectively, that I shouldn't merge recklessly and that I shouldn't dispatch the

book carelessly. But if an objective theory is true, then I ought to do these very things.¹⁷

And this implication, Ross thinks, is counterintuitive. So, I suspect that Ross accepts (2b) because he thinks it's part of commonsense morality that we cannot be obligated to do what we think we shouldn't—cases like the lane-merging case and the book-dispatching case bring this to light. However, as we've seen, an objectivist can plausibly respond to these cases by distinguishing between obligation and praiseworthiness. So, unless Ross has some other reason for accepting the premise (and it's unclear what that would be), we should conclude that his justification for it is weak.¹⁸

It's also unclear why Ross thinks we should accept (2a) and (2c). Here it's more difficult to reconstruct his reasoning. Perhaps Ross is thinking that the principles embodied in these premises are like the principle that “ought” implies “can”: in other words, they are intuitively plausible, fundamental truths about the nature of morality. However, I doubt many will agree. I suspect that many will find these principles to have little intuitive backing and will thus feel no pressure to accept them.

Finally, if these arguments are sound, then they will undermine not only many objective theories, but also many subjective theories, including Ross's own theory. If either R6 or R7 is true, I think it is possible for a person to be obligated to do something

¹⁷ Or so says Ross.

¹⁸ Premise (2b) is similar to one that Ross appeals to in the following argument against objective theories: “It is only by thus distinguishing different rightnesses or suitabilities and by making duty depend on [the subjective sense], that we can do justice to a thought which is inseparable from the thought of duty. This is the thought that anything that we ought to do must be something that we not only can do, but can do with the knowledge or at least the opinion that it is our duty” (1939, 163). Here Ross seems to be assuming that a person cannot be obligated to do something when she does not know, or even believe, that she is obligated to do it. I'd respond to such a claim in the same way that I've responded to (2b).

even though she believes she shouldn't do it. Consider R7. Suppose I have two options, A and B. A uniquely maximizes utility and B is the most morally suitable thing I can do in the circumstances I think I am in, and I know this. Suppose, however, that I am a utilitarian, not a Rossian, and so I believe that I should do A and shouldn't do B. Such a case seems possible.¹⁹ However, if R7 is true, then I am obligated to do something (B) that I believe I shouldn't do. Moreover, since I don't know that I should do B in this case (since I don't even believe that I should do it), this is also a case where I am obligated to do something (B) even though I can't do it because I am obligated to do it—at least if we accept Ross's contention that in order to do something because it is obligatory, one must know that it is obligatory. So, R7 has many of the same (allegedly) unpalatable implications that Ross thinks objective theories have. Similar things apply, I believe, to R6 as well.

¹⁹ One might doubt this. If I were a utilitarian, could I really believe that A uniquely maximizes utility but isn't the most morally suitable thing I could do in the circumstances I believe I'm in? I think so. Perhaps I hold that what I ought to do is not what is the most morally suitable thing I can do in my perceived circumstances, but rather, what is *least* morally *unsuitable* in those circumstances, and I think that action A is least morally unsuitable but isn't most morally suitable. (I will discuss a version of this type of view in chap. 5.) Or maybe I hold that what I ought to do is a function of not only what is morally suitable for me to do in my perceived circumstances, but also a function of what is prudentially suitable for me to do in those circumstances, and I think that the combined function of moral and prudential suitabilities dictates that I do A even though A fails to be most morally suitable. (I will discuss a version of this type of view in chap. 6.) Or maybe I hold that what I ought to do is not what is the most morally suitable thing I can do in the circumstances I think I'm in, but rather what is the most morally suitable thing I can do in the circumstances I think I *should* be in, and I believe that action A is the most morally suitable thing for me to do in these ideal circumstances even though it's not the most morally suitable thing for me to do in my perceived circumstances. In sum, there are lots of ways of filling in the details of the case at hand to make it appear possible.

4.7. Conclusion

I have now examined each of Ross's arguments for abandoning the objective theory of obligation he held in *R&G* and adopting a subjective theory such as R6 or R7 instead. If I have done what I set out to do, then these arguments should seem unimpressive. Many of these arguments appeal to contentious moral principles that Ross makes little attempt to support. There is thus nothing to stop one from rejecting them. Many of Ross's arguments also contain premises that are unwarrantedly strong—it's difficult to see how Ross could justify them. Ross's arguments therefore often need to be weakened in order to have any chance of success. Finally, it's often the case that Ross's arguments, if sound, would cast doubt on not only objective theories of obligation, but also on many subjective theories. In particular, Ross's arguments, if sound, often refute the very theory he ultimately endorses.

CHAPTER 5

ROSSIAN MINIMALISM VS. WORLD ROSSIANISM

5.1. Introduction

In a recent paper, Ned Markosian defends a view he calls “Rossian Minimalism.” According to Rossian Minimalism, an action is morally permissible iff it minimizes total prima facie wrongness. Markosian argues that it is “the best ethical theory that can be stated in terms of Ross’s notion of a prima facie duty” (2009, 9). But I am not convinced. The ethical theory that Ross proposes in *R&G* (namely, RTO_{RG} - RTW_{RG}) seems to me to be more plausible, at least after we transform the theory into a “world” theory. In this chapter, I explain why.

5.2. Some Alternatives to Rossian Minimalism

In his paper, Markosian argues that several moral theories that make use of Ross’s concept of a prima facie duty are problematic. He claims that Rossian Minimalism solves these problems. He concludes that Rossian Minimalism is the best ethical theory available that employs the notion of a prima facie duty. It will be useful for my purposes to retrace the main moves of this dialectic in some detail.

The first ethical theory involving the concept of a prima facie duty that Markosian considers can be stated as follows:

R8: An action A is morally permissible for S(A) to perform if and only if S(A) has a prima facie duty to perform A that is at least as strong as any prima facie duty S(A) has to perform any of its alternatives.^{1, 2}

However, Markosian (2009, 3) argues that R8 is flawed. Suppose I have no prima facie duties to perform any of my alternatives. According to R8, none of my alternatives are morally permissible in this case. But that seems implausible. For any situation I face, there is surely at least one action of mine that is permissible to perform.

However, we can easily modify R8 to get around this problem:

R9: An action A is morally permissible for S(A) to perform if and only if it's not the case that there is an alternative to it, B, such that S(A) has a prima facie duty to perform B that stronger than any prima facie duty S(A) has to perform A.³

R9 solves R8's problem: in a case where I have no prima facie duties to perform any of my alternatives, R9 implies that each of my alternatives is permissible since each

¹ Markosian states R8 as follows: "An act is morally right iff it satisfies a prima facie duty and no alternative satisfies a more stringent prima facie duty" (2009, 3). If, when Markosian speaks of an alternative "satisfying" a prima facie duty, he means that the alternative's agent has a prima facie duty to perform it, then our formulations are equivalent. I will assume throughout this chapter that this is what Markosian means. I, however, prefer not to talk of alternatives "satisfying" prima facie duties. Instead, I prefer to talk of people *having* prima facie duties to perform actions (or to talk of *there being* a prima facie duty for people to perform actions). I prefer to talk this way for two reasons. First, Ross often speaks of a person "having" a prima facie duty to do something (or of "there being" a prima facie duty for a person to do something), but he doesn't speak of an action "satisfying" a prima facie duty. Second, like Markosian, I take prima facie duties to simply be moral reasons (see sec. 1.3). Because of this, I prefer to use "prima facie duty" and "moral reason" more or less interchangeably. And while it makes sense to say that a person has a moral reason to perform an action (or that there is a moral reason for a person to perform an action), it sounds odd to say that a moral reason is satisfied.

² R8 is similar to R3 (see sec. 1.4), though R8 is a theory of permissibility, not obligation.

³ Markosian states R9 as follows: "An act, A, is morally right iff it's not the case that there is an alternative, B, such that B satisfies a more stringent prima facie duty any prima facie duty satisfied by A" (2009, 3).

alternative is such that there is no alternative to it that I have a stronger prima facie duty to perform. However, Markosian (2009, 3) argues that R9 is subject to a different type of problem. Suppose that option A and option B are my only alternatives. I have a prima facie duty to perform A of stringency 5. On the other hand, I have two prima facie duties to perform B, each of stringency 4. According to R9, it is permissible for me to do A and not permissible for me to do B. But that seems incorrect (assuming that everything else about my options is equal).

In response to this, we might alter R9 as follows:

R10: An action A is morally permissible for S(A) to perform if and only if A's total prima facie rightness is at least as great as that of any of its alternatives,

where

The total prima facie rightness of an action A = the sum of the strengths of all of the prima facie duties S(A) has to perform A.^{4,5}

While R10 yields the intuitively correct implications about the cases that trouble R8 and R9, it has other problems. Suppose I can perform either option C or option D. I don't have any prima facie duties *to* perform either option. However, I have a prima facie duty *not* to perform C of stringency 5. If everything else about my options is equal, R10

⁴ Markosian states R10 as follows: "An act is morally right iff it maximizes PFD utility," where "The PFD utility of act A =df. the sum of the strengths of all the prima facie duties that A's agent would satisfy by performing A" (2009, 4). Markosian calls R10 "Rossian Utilitarianism."

⁵ This is equivalent to the way I defined "total prima facie rightness" earlier (see sec. 1.4). Later, I will modify this definition slightly by adding some time indexes to it (see note 12). For now, these indexes aren't important.

implies that each is permissible. But this seems counterintuitive: surely I am not permitted to perform C in this case.

The solution to this problem seems clear. We should modify R10 to

R11: An action A is morally permissible for S(A) to perform if and only if A's total prima facie rightness minus its total prima facie wrongness is at least as great as that of any of its alternatives,

where

The total prima facie wrongness of an action A = the sum of the strengths of all of the prima facie duties S(A) has *not* to perform A.⁶

Notice that R11 is identical to Ross's theory of permissibility in *R&G* (at least if we assume that S(A) can perform A and that the right-hand side of R11 specifies what makes A permissible). R11 (i.e., RTP_{RG}) is much more plausible than R8-R10; it solves all of the problems discussed so far.

Markosian, however, raises an insightful objection to RTP_{RG}.⁷ Suppose I have promised my brother that I will mow his lawn by the end of the day. It's now noon and I can either mow his lawn or eat lunch. (If I eat lunch now, I will mow his lawn after lunch.) Mowing my brother's lawn now will keep my promise to him. Presumably, I therefore have a prima facie duty to do it. On the other hand, eating lunch now won't

⁶ I will also later modify this definition by adding some time indexes to it (see note 12).

⁷ Markosian actually raises this objection to R10, not RTP_{RG}. Markosian never discusses RTP_{RG} in his paper. This is odd since, we've seen, RTP_{RG} is the theory of permissibility that Ross actually seems to accept (at least in *R&G*), and it seems, at first glance, to be the most plausible way of developing an ethical theory that makes use of the notion of a prima facie duty that solves the problems Markosian has raised thus far. However, since the objection Markosian raises to R10 applies just as well to RTP_{RG}, I will proceed as if he objects to RTP_{RG}.

break my promise. It's therefore not the case that I have a prima facie duty not to eat lunch now in virtue of its being a promise breaking. Suppose, in fact, that I have no prima facie duties to (or not to) eat lunch now. If that's the case (and everything else is equal), then RTP_{RG} implies that I am permitted to mow my brother's lawn now, but I am *not* permitted to eat lunch now (and mow the lawn later). But this is counterintuitive; surely each of my options is permissible in this case. Markosian concludes: "The problem, in short, is that $[RTP_{RG}]$ requires one to satisfy every prima facie duty *as soon as possible*, whereas in truth it is often permissible to procrastinate a little bit before doing one's duty. I will call this The Procrastination Problem" (2009, 4).

As Markosian suggests (2009, 4-5), a promising way of responding to this problem is to transform RTP_{RG} into a "world" theory. However, before exploring what this transformation would look like, I would like to discuss another important problem for RTP_{RG} , one that will also push a proponent of the theory to adopt a "world" version of it.

5.3. The Bergström/Castañeda Problem

Back in the 1960s, Lars Bergström (1966, 33-36) and Hector-Neri Castañeda (1968) raised a perplexing problem for act utilitarianism.⁸ However, the problem also affects many other moral theories, including RTO_{RG} - RTW_{RG} . When applied to RTO_{RG} - RTW_{RG} , the problem is this. Recall RTO_{RG} :

RTO_{RG} : An action A is morally obligatory for S(A) to perform if and only if (and because) S(A) can perform A, and A's total prima facie rightness minus its total prima facie wrongness is greater than that of any other action S(A) can perform instead.

⁸ This problem is nicely summarized in Feldman (1986, chap. 1).

For ease of exposition, it will be useful at this point to state RTO_{RG} in a somewhat more perspicuous way:

RTO_{RG} : An action A is morally obligatory for S(A) to perform if and only if (and because) A uniquely maximizes net prima facie rightness,

where

An action A uniquely maximizes net prima facie rightness =_{df.} S(A) can perform A, and A's total prima facie rightness minus its total prima facie wrongness is greater than that of any other action S(A) can perform instead.⁹

Now suppose that I can perform A, B, or A and B (that is, I can perform the “compound” action A + B). (For instance, suppose I am at a concert and, at the end of the concert, I can either stand up (A), clap (B), or stand up and clap (A + B)). Suppose also that I am obligated to perform A + B. If I am obligated to perform A + B, then, presumably, I am also obligated to perform A. Intuitively, obligation distributes through conjunction. But a proponent of RTO_{RG} won't be able to say that I'm obligated to perform A in this case. By stipulation, I'm obligated to perform A + B. So according to RTO_{RG} , it uniquely

⁹ To say that an action maximizes net prima facie rightness (but not necessarily *uniquely*) is to say that S(A) can perform A, and A's total prima facie rightness minus its total prima facie wrongness is at least as great as that of any other action S(A) can perform instead. Given this, we can also state RTP_{RG} and RTW_{RG} using “maximizing” language:

RTP_{RG} : An action A is morally permissible for S(A) to perform if and only if (and because) A maximizes net prima facie rightness;

RTW_{RG} : An action A is morally wrong for S(A) to perform if and only if (and because) S(A) can do A, and A fails to maximize net prima facie rightness.

The careful reader may have noticed that I occasionally used the term “maximizes” in previous chapters, but I used it inconsistently: sometimes I used it in roughly the same way that I'm using it here, but other times I used it to mean what I here mean by “uniquely maximizes.” I hope this hasn't caused much confusion.

maximizes net prima facie rightness. But if A + B uniquely maximizes net prima facie rightness, then A *doesn't* uniquely maximize net prima facie rightness (by the meaning of “unique”). And if A doesn't uniquely maximize net prima facie rightness, then according to RTO_{RG} , I'm not obligated to do it.

Here's another, related problem. Suppose I can either perform A, or perform B. Suppose, however, that I can't perform A without performing B. (For instance, suppose I can pick up my brother at the airport (A), or I can get into my car (B), but I can't pick up my brother at the airport unless I first get into my car and drive there.) Suppose also that I am obligated to perform A. If I am obligated to perform A, then I am surely also obligated to perform B. After all, performing B is a necessary prerequisite of performing A, and, intuitively, if I am obligated to perform an action, then I am also obligated to perform any necessary prerequisite of it. But a proponent of RTO_{RG} can't say that I am obligated to perform B in this case. By stipulation, I am obligated to perform A. According to RTO_{RG} , it therefore uniquely maximizes net prima facie rightness. But if A uniquely maximizes net prima facie rightness, then B doesn't. And if B doesn't uniquely maximize net prima facie rightness, then according to RTO_{RG} , it's not obligatory.

There are several ways one might try to respond to these puzzles. First, one might deny that we can perform conjunctive actions. This will solve the first problem since it clearly relies on our being able to perform conjunctive acts. It may also solve the second problem since it involves an action that has a necessary prerequisite, and one might argue that any such action will be a conjunctive one. However, denying that we can perform conjunctive acts is a highly revisionary proposal. It certainly seems like we can perform them. We can do things like type and listen to music at the same time, raise our right

hand and our left hand at the same time, and walk and think at the same time (indeed, I've done each of these things today). But these appear to be a conjunctive acts. Moreover, I can surely do complex things such as ride my bike home tonight. However, riding my bike home tonight involves unlocking my bike from the bike rack, putting on my helmet, getting on my bike, pedaling for a while, turning left at the stop sign, getting off my bike, bringing my bike into my apartment, etc. In other words, riding my bike home tonight involves performing a conjunctive act. So, if I can ride my bike home tonight (which I surely can), then it seems that I can perform a conjunctive act. Denying that we can perform conjunctive acts is therefore implausible.

Another way of responding to the Bergström/Castañeda puzzles is to deny that “ought” distributes through conjunction and deny that we are obligated to perform any necessary prerequisite of the things we are obligated to do. But this is also an unattractive solution. These principles are, after all, highly intuitive. It would be nice to find a solution to the puzzles that leaves these principles intact.

A third way to respond is to deny that A and $A + B$ are alternatives to each other in the first puzzle and that A and B are alternatives to each other in the second. This is the solution adopted by Bergström (1966, chap. 2). He claims that a set of actions is a set of alternatives for a person if and only if the person can perform each of the actions in the set, and the members of the set are time identical (i.e., they take place at the same time), mutually exclusive (i.e., the person can't perform more than one of them), and jointly exhaustive (i.e., the person must perform one of them). On this conception of alternatives, $\{A, B, A + B\}$ is not an alternative set for me in the first puzzle since its members are neither time identical nor mutually exclusive. A proponent of RTO_{RG} is

therefore not committed to denying that A is obligatory if $A + B$ is. For the same reasons, $\{A, B\}$ is not an alternative set for me in the second puzzle, and so a proponent of RTO_{RG} is not committed to denying that B is obligatory if A is.

Unfortunately, as Bergström recognizes, this solution has a serious problem. On Bergström's conception of alternatives, there can be multiple alternative sets for a person at a time. Suppose, for instance, that I can now perform A, $\sim A$, $A + B$, or $\sim A + B$. Suppose that A and $\sim A$ are time identical, mutually exclusive, and jointly exhaustive. The set $\{A, \sim A\}$ is therefore an alternative set for me. But suppose that $A + B$ and $\sim A + B$ are also time identical, mutually exclusive, and jointly exhaustive. The set $\{A + B, \sim A + B\}$ is therefore also an alternative set for me. But now suppose that the net prima facie rightness of these actions is as follows:

Table 3: A problem for Bergström's conception of alternatives

<u>Alternative</u>	<u>Net Prima Facie Rightness</u>
A	10
$\sim A$	5
$A + B$	15
$\sim A + B$	20

If a proponent of RTO_{RG} accepts Bergström's conception of alternatives, then she'll be committed to saying that, relative to the alternative set $\{A, \sim A\}$, I am obligated to perform A, and relative to the alternative set $\{A + B, \sim A + B\}$, I am obligated to perform $\sim A + B$. But these prescriptions are incompatible: I cannot perform both A and $\sim A + B$.¹⁰

¹⁰ One might wonder whether the scenario I've described is possible. I think it is. To see this, it may be helpful to make the case a bit more concrete. Suppose I'm in a room and I have two buttons in front of me, a green one and a red one. There are four things I can do: press the green

So, if a proponent of RTO_{RG} adopts Bergström’s conception of alternatives, she’ll have to say that only *one* of a person’s alternative sets determines her obligations. But then she’ll need to provide some criterion for determining which set this is. Bergström (1966, chap. 2) attempts to do just that, but he admits that none of his proposals is wholly satisfactory. A proponent of RTO_{RG} might try to do better. However, it seems to me more promising for her to abandon Bergström’s solution to his puzzle altogether and try to solve it in a different way: by transforming her view into a “world” view. In the next section, I’ll describe what such a view would look like.

5.4. World Rossianism

There are several ways of transforming RTO_{RG} into a “world” view. Feldman (1975, 1986) and Smith (1976, 1978) each develop “world” versions of utilitarianism. The “world” version of RTO_{RG} that I will present here (I’ll call it “ $WRTO_{RG}$ ”) incorporates elements of each of these accounts.

In order to state $WRTO_{RG}$, I first need to say what it is for a world to be “accessible” to a person at a time. The rough idea is this (I will consign the technical details to a footnote). Consider all of the ways you could now live out your life. For

button now (A), refrain from pressing it now ($\sim A$), press it now and press the red button later ($A + B$), or refrain from pressing the green button now and press the red one later ($\sim A + B$). If I press the green button now, someone will get a 10 unit jolt of pleasure now. If I refrain from pressing the green button now, someone will get a 5 unit jolt of pleasure now. However, if I press the green button now and the red one later, then someone will get a 10 unit jolt of pleasure now and a 5 unit jolt later. But if I refrain from pressing the green button now and press the red one later, then someone will get a 5 unit jolt now and a 15 unit jolt later. It seems possible for me to be in a scenario such as this, strange though it is. It also seems possible that the above chart correctly indicates the net *prima facie* rightness of my options in the scenario.

each of them, consider the possible world that would be actual if you were to live out your life in that way. These are the possible worlds that are accessible to you now. As time goes by, there will be fewer and fewer ways you could live out your life, and thus fewer and fewer possible worlds accessible to you. Eventually, when you are very near death, there may be only one world accessible to you then. But early on in life, it's likely that there are a tremendous number of possible worlds accessible to you.¹¹

¹¹ Here is a more technical explanation of what it is for a world to be accessible to a person at a time. First, I will take as a primitive the concept of a person being able, at a time, to perform an action at that time. With this concept in hand, I can say what a "sequence" of actions for a person at a time is:

$\{A_1, A_2, A_3, \dots, A_n\}$ is a sequence of actions for a person S at a time $t =_{df.}$ S can at t perform A_1 at t , and if S were to do that, then she would be able at $t(A_2)$ to perform A_2 at $t(A_2)$, and if she were to do that, then she would be able at $t(A_3)$ to perform A_3 at $t(A_3)$, and...and if she were to do that, then she would be able at $t(A_n)$ to perform A_n at $t(A_n)$,

where

$t(A) =_{df.}$ the time of action A .

There are different kinds of sequences of actions. Sequences of actions can be maximal or non-maximal. A maximal sequence of actions can be understood as follows:

- $\{A_1, A_2, A_3, \dots, A_n\}$ is a maximal sequence of actions for a person S at a time $t =_{df.}$
- (a) $\{A_1, A_2, A_3, \dots, A_n\}$ is a sequence of actions for S at t , and
 - (b) There is no sequence of actions for S at t that contains each member of $\{A_1, A_2, A_3, \dots, A_n\}$ plus an (some) additional member(s).

(A non-maximal sequence of actions is simply a sequence of actions that is not maximal, in the sense just defined.) With that, I can now say what it is for a world to be accessible to a person at a time:

A world W is accessible to a person S at a time $t =_{df.}$ there is a maximal sequence of actions for S at t such that if S were to perform each member of the sequence at its respective time, W would be actual.

To state $WRTO_{RG}$, I also need to say what the net prima facie rightness of a world is for a person at a time. The general idea is this (again, I'll leave the technical details for a footnote). Consider a world that is accessible to you now. Now consider each of the things you do in that world either now or in the future. For each of these things, calculate its total prima facie rightness minus its total prima facie wrongness (that is, for each of these things, determine the total strength of the prima facie duties you have at its time to perform it then, and subtract from this quantity the total strength of the prima facie duties you have at its time *not* to perform it then). Let's call the total prima facie rightness minus the total prima facie wrongness of an action its "net prima facie rightness." Add together the net prima facie rightness of each of the things you do in the world (now or in the future). This gives us the net prima facie rightness of the world for you now.¹²

One more preliminary task: I need to explain what it is for a world to "maximize" net prima facie rightness. The idea here is simple. To say that a world maximizes net prima facie rightness for you now is to say that it's accessible to you now, and its net

¹² Here are the technical details:

The total prima facie rightness of an action A = the sum of the strengths of all of the prima facie duties S(A) has at t(A) to perform A at t(A). (Note: this is one of the modifications I earlier promised to make (see note 5).)

The total prima facie wrongness of an action A = the sum of the strengths of all of the prima facie duties S(A) has at t(A) *not* to perform A at t(A). (Note: this is the other modification I earlier promised to make (see note 6).)

The net prima facie rightness of an action A = (the total prima facie rightness of A) – (the total prima facie wrongness of A).

The net prima facie rightness of a world W for a person S at a time t = the sum of the net prima facie rightness of each of the actions S performs in W at t or after t.

prima facie rightness (for you now) is at least as great as that of any other world accessible to you now.¹³

With that, I can now finally state $WRTO_{RG}$. The general idea is this: a person is obligated to do something just in case she does it in each world accessible to her that maximizes net prima facie rightness. More precisely,

$WRTO_{RG}$: It is morally obligatory, as of a time t , for $S(A)$ to perform an action A at $t(A)$ if and only if (and because) $S(A)$ performs A at $t(A)$ in each world that maximizes net prima facie rightness for her at t .¹⁴

I can now also state “world” versions of RTP_{RG} and RTW_{RG} :

$WRTP_{RG}$: It is morally permissible, as of a time t , for $S(A)$ to perform an action A at $t(A)$ if and only if (and because) $S(A)$ performs A at $t(A)$ in some world that maximizes net prima facie rightness for her at t ;

$WRTW_{RG}$: It is morally wrong, as of a time t , for $S(A)$ to perform an action A at $t(A)$ if and only if (and because) $S(A)$ performs A at $t(A)$ in a world accessible to her at t ,¹⁵ but she performs it in no world that maximizes net prima facie rightness for her at t .

¹³ Slightly more formally:

A world W maximizes net prima facie rightness for a person S at a time $t =_{df}$.

(a) W is accessible to S at t , and

(b) W 's net prima facie rightness for S at t is at least as great as that of any other world accessible to S at t .

¹⁴ Even this may need some tinkering if we admit that a person can be in a situation where an infinite number of worlds are accessible to her, and for each one, there is some world accessible to her whose net prima facie rightness is greater. I will ignore this kind of case— $WRTO_{RG}$ is complicated enough already. For more on this sort of case and how $WRTO_{RG}$ might be modified to accommodate it, see Feldman (1986, 37-38).

¹⁵ This clause is here to ensure that “wrong” implies “can.”

I will use the term “World Rossianism” to refer to the combination of $WRTOR_G$, $WRTPR_G$, and $WRTWR_G$.

World Rossianism nicely handles the Bergström/Castañeda problem. Any world where I perform A + B is also a world where I perform A. So, if I am obligated to perform A + B, World Rossianism implies that I am also obligated to perform A. Likewise, if performing B is a necessary prerequisite of performing A, then any world where I perform A is also one where I perform B. So, if I am obligated to perform A, then World Rossianism implies that I’m also obligated to perform B. World Rossianism also solves the problems that Markosian raises to R8-R11, including the Procrastination Problem. If we assume, as seems plausible, that a world where I mow my brother’s lawn now and eat lunch later maximizes net prima facie rightness for me now, as does a world where I eat lunch now and mow my brother’s lawn later, then according to World Rossianism, mowing my brother’s lawn now and eating lunch now are each morally permissible for me now, just as it seems. (I trust that the reader can easily see how World Rossianism solves the other problems that Markosian raises to R8-R11.)

While World Rossianism seems to be the best moral theory that makes use of the notion of a prima facie duty that we have encountered so far, Markosian thinks it is not perfect—he thinks it is subject to two serious objections. In the next section, I will discuss these objections. I will argue, however, that they can be overcome.¹⁶

¹⁶ Markosian actually objects to a “world” version of R10, not to World Rossianism precisely as I’ve formulated it. However, the objections he raises to a “world” version of R10 apply just as well to World Rossianism. I’ll thus proceed as if he objects to it.

5.5. Problems for World Rossianism

Here's the first objection that Markosian raises to World Rossianism. Suppose there are only two worlds accessible to me now.¹⁷ In one of them, I live out the rest of my life by performing actions that are, for the most part, neutral with respect to prima facie duty. That is, I live out the rest of my life by performing actions that I have very little prima facie duty to (or not to) perform. In this world, I spend the majority of my life by myself, engaged in some very dull activity. In the other world accessible to me now, I live out the rest of my life by making a constant barrage of trivial promises (such as the promise to continue breathing) and then keeping them. If we assume that there is a prima facie duty of promise keeping, then the net prima facie rightness of the trivial-promise-keeping world will be quite large. On the other hand, the net prima facie rightness of the dull-activity world will be close to 0. World Rossianism therefore implies that I am morally obligated to live out my life by making and keeping a multitude of trivial promises. But that seems counterintuitive. Markosian calls this objection the "Gratuitous Duty Objection" (2009, 5-6).

There are several ways for a World Rossian to respond to this objection. First, she could deny that there is a prima facie duty of promise keeping. If there is no prima facie duty of promise keeping, then it's not at all clear that the net prima facie rightness of the trivial-promise-keeping world is greater than that of the dull-activity world, and it's thus not clear that World Rossian implies that I should spend my life making and keeping trivial promises.

¹⁷ Markosian makes this assumption to keep the case that follows relatively simple. Of course, in almost any actual scenario, there will be far more than two worlds accessible to me.

One might object that this response is *ad hoc*. However, there may be some independent reasons for denying that there is a prima facie duty of promise keeping. Holly Smith (1997), for instance, attempts to provide such a reason. She has us consider the following case. Suppose I'm a dean at a university, and I have to decide whether to fund a new faculty position in either the Classics Department or the History Department. History could use the new faculty position more, and giving the position to History would be best for the university. However, if I give the position to Classics, I know that my cousin, with whom I am very close, will get the job. I'd love for this to happen, but I'd also like to do the right thing. If there were a prima facie duty of promise keeping, then I could easily *make* it permissible for me to give the position to Classics. I could do this by making a (or several) solemn promise(s) to give the position to Classics. But surely I can't manipulate morality in this way. So, Smith concludes, we should deny that there is a prima facie duty of promise keeping.

There may be other reasons for thinking that there is no prima facie duty to keep one's promises. Suppose I make a promise to do something terrible, such as a promise to murder someone. My murdering someone keeps a promise I've made, but one might think that I have no prima facie duty to do this. Or suppose a thief breaks into my house. Noticing suspicious activity at my home, a police officer comes to my door. The thief threatens to kill me unless I promise to tell the officer that nothing is wrong. I immediately make the promise. My telling the officer that nothing is wrong fulfills my promise to the thief, but, one might think, I have no prima facie duty to do this.

Of course, all of these arguments are controversial. But my goal here is not to come to any firm conclusion about their soundness. I'm merely trying to point out that if a

World Rossian responds to the Gratuitous Duty Objection by denying that there is a prima facie duty of promise keeping, she may be able to circumvent the charge that her denial is *ad hoc* by offering independent reasons to support her denial—in fact, there seem to be independent reasons that she could readily appeal to, reasons that are at least prima facie plausible.

However, it seems to me that a World Rossian could also respond to the Gratuitous Duty Objection in a somewhat different way. She could claim that there is, in general, a prima facie duty of promise keeping, but only *some* promise keepings contribute to the net prima facie rightness of a world. In particular, she could say that the net prima facie rightness of a world for a person at a time is only boosted by the promise keepings in it that keep promises she's made *prior* to that time. So, suppose it's noon on Feb. 29th, and I'm trying to decide what to do. In one of the worlds accessible to me at this time, I make a promise ten years from now, and keep it shortly thereafter. According to the present suggestion, the net prima facie rightness of this world for me at noon on Feb. 29th is not increased by containing this promise keeping. This is because this promise keeping is the keeping of a promise I haven't yet made.

Accepting this proposal would allow a World Rossian to solve the Gratuitous Duty Objection, for the world where I make and keep a bunch of trivial promises is filled with promise keepings that keep promises I haven't yet made. They will thus not contribute to the net prima facie rightness of this world for me now, and its net prima facie rightness for me now is no longer clearly greater than that of the world filled with dull activity.

One might think that this proposal only temporarily solves the problem. Suppose we modify the example so that I've *already* made the multitude of trivial promises. I

now have only two worlds accessible to me: one where I spend the rest of my life fulfilling the trivial promises I've made, and one where I spend the rest of my life performing actions that are neutral with regard to prima facie duty. Even if a World Rossian accepts the present proposal, it seems she'll have to admit that the net prima facie rightness of the former world for me now is very high, since it contains promise keepings that keep promises I've already made. She'll then be forced to admit that I ought to live out my life keeping the promises.

But I don't find this counterintuitive. I probably shouldn't have made all of those trivial promises in the first place. At many points in my past, it probably wasn't the case that I made a trivial promise in any world that maximized net prima facie rightness. But now that I've made those promises, it seems to me that I'm obliged to keep them, at least if my only other alternative is to live a dull, prima-facie-duty-neutral life.

So much for the Gratuitous Duty Objection. The second objection that Markosian raises to World Rossianism is this. Suppose I'm walking on a path in a park and I come to a fork in the path. At that point, there are only two worlds accessible to me: one where I take the left side of the path, and one where I take the right.¹⁸ Suppose that in the world where I bear left, I am ambushed by a mugger and tragically murdered. On the other hand, in the world where I bear right, I enjoy a nice walk through the park, arrive home safely, and live a long and happy life. Suppose that the net prima facie rightness of the world where I take the right path is greater than that of the world where I take the left. World Rossianism then implies that I'm obligated to go right; it's wrong for me to go

¹⁸ Again, this is a simplifying assumption.

left. But this is implausible. After all, I have no idea that one side of the path will lead to my demise whereas the other will allow me to flourish. Given this, it is surely permissible for me to take the left path. Markosian calls this objection the “Walk in the Park Objection” (2006, 6).

I think there are several ways for a World Rossian to respond to this objection. First, she can appeal to the objective/subjective distinction I made in the previous chapter. She can then agree with Markosian that in the Walk in the Park scenario, it’s permissible for me to take the path that leads to my murder. However, she can insist that the sense of “permissible” according to which it’s permissible for me to take this path is the *subjective* sense. But, she can go on to say, World Rossianism is a theory of the objective sense of “obligation,” “permissibility,” and “wrongness,” not a theory of the subjective sense of these terms. It therefore doesn’t conflict with the thought that it’s permissible, subjectively speaking, for me to take the left, unsafe path. Instead, it implies that I ought, objectively speaking, to take the path that leads to my survival. But this, a World Rossian can claim, is in no way problematic.

However, a World Rossian can respond to the Walk in the Park Objection without having to maintain that there is an objective and a subjective sense of “ought,” “permissible,” and “wrong.” Instead, she can claim that these terms are univocal and respond to the objection in one of two ways. First, she can leave her theory as it is and deny that it makes counterintuitive conclusions about the Walk in the Park scenario. In other words, she might insist that I am not in fact permitted to take the path that leads to my being murdered, just as World Rossianism implies. Perhaps I would be *blameless* for taking this path, but, a World Rossian might maintain, it’s nonetheless wrong for me to

do so. Markosian (2009, 6) seems to find this type of response is unconvincing, but he doesn't elaborate on the matter much. A World Rossian might do well to press him on the issue.

On the other hand, if a World Rossian doesn't want to distinguish between an objective and a subjective sense of "ought"/"permissible"/"wrong," and she also doesn't want to deny that I am permitted to take the unsafe path, she can respond to the objection in another way: she can simply "subjectivize" her theory in roughly the way that Ross does in *FE*. In particular, she can maintain that the obligatoriness/permissibility/wrongness of an action is determined, at least in part, by the agent's mental state. Consider, for instance, the following version of World Rossianism:

Subjective World Rossianism

It is morally obligatory, as of a time *t*, for *S(A)* to perform an action *A* at *t(A)* if and only if (and because) *S(A)* *believes at t* that she performs *A* at *t(A)* in each world that maximizes net prima facie rightness for her at *t*;

It is morally permissible, as of a time *t*, for *S(A)* to perform an action *A* at *t(A)* if and only if (and because) *S(A)* *believes at t* that she performs *A* at *t(A)* in some world that maximizes net prima facie rightness for her at *t*;

It is morally wrong, as of a time *t*, for *S(A)* to perform an action *A* at *t(A)* if and only if (and because) *S(A)* *believes at t* that she performs *A* at *t(A)* in a world accessible to her at *t*, but she performs it in no world that maximizes net prima facie rightness for her at *t*.

Or consider a version of World Rossian that is just like Subjective World Rossianism except that each instance of "believes" is replaced with "is justified in believing." Each of these modifications of World Rossianism generates the intended result, at least if I believe (and am justified in believing) that I take the left (unsafe) path in a world that maximizes net prima facie rightness.

Let me be clear: I don't mean to suggest that either of the modifications to World Rossianism suggested in the previous paragraph is ultimately tenable. Indeed, I think that each has serious problems.¹⁹ However, I do suspect that something like them can be made to work. So, I think that even if a World Rossian holds that "obligation," "permissibility," and "wrongness" are univocal, and she holds that I am permitted to take the unsafe path, she can still respond to the Walk in the Park Objection without abandoning her theory completely and accepting something else, like Rossian Minimalism. She can respond to the objection simply by subjectivizing her theory and incorporating into it some aspect of the agent's mental state.

For these reasons, I do not think that World Rossianism is as problematic as Markosian suggests. However, even if I'm right about this, it may turn out that Markosian's own theory, Rossian Minimalism, is even *less* problematic than World Rossianism is. In the remainder of this chapter, I will argue that this is not the case. Rossian Minimalism has many of the same difficulties that World Rossianism has, and there are even some reasons for preferring the latter theory to the former. So, "the best ethical theory that can be stated in terms of Ross's notion of a prima facie duty" seems to me to be World Rossianism, not Rossian Minimalism.

5.6. Rossian Minimalism

As I understand it, Rossian Minimalism is the following view:

¹⁹ One of the most serious problems with these theories is that they make the wrong conclusions about Frank Jackson's Dr. Jill case. I'll discuss this case in chapter 7.

An action A is morally permissible for S(A) to perform if and only if (and because) A minimizes prima facie wrongness,

where

An action A minimizes prima facie wrongness =_{df.} S(A) can perform A, and A's total prima facie wrongness is at least as small as that of any other action S(A) can perform instead.²⁰

I will also assume that Rossian Minimalists accept the following claims about obligation and wrongness:

An action A is morally obligatory for S(A) to perform if and only if (and because) A uniquely minimizes prima facie wrongness;

An action A is morally wrong for S(A) to perform if and only if (and because) S(A) can do A, and A fails to minimize prima facie wrongness.

Henceforth, I will use the term “Rossian Minimalism” to refer to the combination of the theories of permissibility, obligation, and wrongness just stated.

Rossian Minimalism starkly contrasts with R10, which says that an action is permissible iff it *maximizes* total prima facie *rightness*. It also contrasts with R11 (RTP_{RG}), which says that an action is permissible iff it maximizes *net* prima facie rightness (over wrongness). In addition, since it is not a “world” theory, Rossian Minimalism contrasts with World Rossianism.

Let's now see how Rossian Minimalism deals with the problems that hinder the other Rossian theories discussed in this chapter. Rossian Minimalism nicely solves many

²⁰ Markosian states Rossian Minimalism as follows: “An act is morally right iff it minimizes prima facie duty violations by its agent” (2009, 7).

of these problems. Consider, for instance, R8. It is problematic because it implies that in a scenario where I have no prima facie duties to (or not to) perform any of my alternatives, none of my alternatives is permissible. Rossian Minimalism implies that in such a case, each of my options is permissible since the total prima facie wrongness of each is 0, and thus each minimizes prima facie wrongness. Consider also R10. It is problematic because it implies that in a situation where I have no prima facie duty to perform either C or D, but a strong prima facie duty not to perform C, it is permissible for me to perform either action. Rossian Minimalism implies that only B is permissible, since only it minimizes prima facie wrongness. Next, recall R11 (RTP_{RG}). It is subject to the Procrastination Problem: in a case where I can either keep my promise to my brother now, or procrastinate a little and keep it later, RTP_{RG} implies that it is not permissible for me to procrastinate. According to Rossian Minimalism, each option is permissible, for neither is prima facie wrong and each therefore minimizes prima facie wrongness. Finally, recall that World Rossianism is subject to the Gratuitous Duty Objection: in a scenario where I can either live out my life by making and keeping a multitude of trivial promises, or live out my life by engaging in prima-facie-duty-neutral activity, World Rossianism implies (allegedly) that I am obliged to do the former. On the other hand, Rossian Minimalism implies that each option is permissible, since the prima facie wrongness of each is presumably 0, and thus each minimizes prima facie wrongness.

So, Rossian Minimalism solves many of the problems that plague the other Rossian views discussed so far. However, it seems to me that it does not solve all of these problems. Consider R9. It is problematic because it implies that in a scenario where I

have a prima facie duty to perform A of stringency 5, and two prima facie duties to perform B of stringency 4, it is permissible for me to perform A and not permissible for me to perform B. However, it looks like this case also causes trouble for Russian Minimalism. Since I have no prima facie duty not to perform either of my options, each minimizes net prima facie wrongness, and Russian Minimalism therefore implies that each is permissible. But that seems incorrect. It seems that I should do B in this case; performing A would be wrong. (I will return to this type of case in the next section.)

Moreover, Russian Minimalism doesn't solve the Bergström/Castañeda Problem. Recall that the Bergström/Castañeda Problem is a problem for RTO_{RG} . In a situation where I am obliged to perform $A + B$, a proponent of RTO_{RG} cannot say that I am also obliged to perform A. Likewise, in a situation where I am obliged to perform A, a proponent of RTO_{RG} cannot say that I am also obliged to perform B, a necessary prerequisite of A. As I've said, these problems seem to push a proponent of RTO_{RG} to accept a "world" version of her theory. However, it seems to me that these problems also hamper Russian Minimalism and push a proponent of it to adopt a "world" version of the view. Let me explain.

Suppose that I have three alternatives: A, B, and $A + B$. Suppose also that $A + B$ is obligatory. If I am obligated to perform $A + B$, then I'm presumably also obligated to perform A. But a Russian Minimalist can't say this. Since I'm obligated to perform $A + B$, it uniquely minimizes prima facie wrongness. But if $A + B$ uniquely minimizes prima facie wrongness, then A doesn't. So, A isn't obligatory. Relatedly, suppose I have two alternatives: A and B. Suppose that A is obligatory and that B is a necessary prerequisite of A. If A is obligatory, then presumably B is as well. But a Russian Minimalist can't

say this. Since A is obligatory, it uniquely minimizes prima facie wrongness. But if it uniquely minimizes prima facie wrongness, then B doesn't. So, B isn't obligatory.

A Rossian Minimalist might try to respond to these problems in the same way that a proponent of RTO_{RG} might try to respond to them: (a) she might deny that we can perform conjunctive actions; (b) she might deny the principle that "ought" distributes through conjunction and the principle that we are obliged to perform any necessary prerequisite of our obligations; (c) she might deny that $\{A, B, A + B\}$ is an alternative set in the first problem and that $\{A, B\}$ is an alternative set in the second; or (d) she might change her view to a "world" view. For the reasons explained above, (d) seems to me to be the most plausible solution.

However, as Markosian notes (2009, 11-12), if a Rossian Minimalist changes her view to a "world" view, she'll then be subject to a version of the Walk in the Park Objection. To see this, consider what a "world" version of Rossian Minimalism (I'll call it "World Minimalism") would look like:

World Minimalism

It is morally obligatory, as of a time t , for $S(A)$ to perform an action A at $t(A)$ if and only if (and because) $S(A)$ performs A at $t(A)$ in each world that minimizes prima facie wrongness for her at t .

It is morally permissible, as of a time t , for $S(A)$ to perform an action A at $t(A)$ if and only if (and because) $S(A)$ performs A at $t(A)$ in some world that minimizes prima facie wrongness for her at t .

It is morally wrong, as of a time t , for $S(A)$ to perform an action A at $t(A)$ if and only if (and because) $S(A)$ performs A at $t(A)$ in a world accessible to her at t , but she performs it in no world that minimizes prima facie wrongness for her at t .²¹

²¹ Here are the definitions of the new terms involved in World Minimalism:

Now suppose I'm walking through a park and I come to a fork. I then have only two worlds accessible to me. In one, I take the right-hand side, make it home safely, and the rest of my life is happy and largely free of actions that I have a prima facie duty not to perform. In the other world accessible to me, I take the left-hand side, which leads to my being brutally murdered. Let's assume that the world where I take the right side minimizes prima facie wrongness, but the world where I take the left side does not.²² World Minimalism then implies that I am obligated to take the safe path; it's wrong for me to take the tragic path. But this may seem implausible since I have no clue about the prima facie wrongness of the worlds accessible to me.

A World Minimalist can respond to this objection in the same way that a World Rossian can. First, she can say that it's only true that I am morally permitted to take the tragic path when we have the subjective sense of "permitted" in mind; however, World Minimalism is a theory of the objective sense of "permitted," and so it doesn't conflict with this thought. On the other hand, if a World Minimalist prefers not to posit the

The prima facie wrongness of a world W for a person S at a time t = the sum of the total prima facie wrongness of all of the actions S performs in W at t or after t;

A world W minimizes prima facie wrongness for a person S at a time t =_{df.}

- (a) W is accessible to S at t, and
- (b) W's prima facie wrongness for S at t is at least as small as that of any other world accessible to S at t.

²² I'm assuming here that there is a prima facie duty not to cause harm to oneself. If this assumption is false, then the world where I take the path that leads to my being murdered may minimize prima facie wrongness after all. However, I could then make the same point I'm trying to make by altering the scenario so that it's not me, but someone else, who is murdered if I take the left path, but no one is harmed if I take the right. In this case, the world where I take the left path would not minimize prima facie wrongness, for there is surely a prima facie duty not to harm others.

existence of two senses of “obligatory”/“permissible”/“wrong,” then she can either deny the intuition that I am permitted to take the path that leads to my demise, or she can accept the intuition and subjectivize her moral theory so that it is no longer inconsistent with it. There are thus several plausible ways for a World Minimalist to respond to the Walk in the Park Objection; however, the central point I’m trying to make is that the problem (if it is one) is one that affects World Minimalism just as much as it affects World Rossianism.

In fact, I think a version of the Walk in the Park objection arises even for the original, “non-world” version of Rossian Minimalism. Markosian suggests that the objection hinders only “world” views, such as World Rossianism and World Minimalism, but I think this is false.

Suppose again that I’m at a fork in a path; either I can veer left, or I can veer right. Suppose I have no prima facie duty not to go right. However, if I go left, I will get murdered, and suppose I’ve made a (rather unusual) promise to my wife not to do things that will get me murdered. Going on the left path breaks this promise, and I therefore (presumably) have a prima facie duty not to do it. In this case, Rossian Minimalism implies that I’m permitted to go right, but I’m not permitted to go left. But if one finds it counterintuitive to deny that it’s permissible for me to veer left in the original Walk in the Park case, then one should also find it counterintuitive to deny that it’s permissible for me to veer left in the new version of the case. After all, in the new case I still have no clue that going left will lead to my being murdered, and I therefore have no clue that doing so will break my promise to my wife. But considerations like these are the very sort of thing

that motivates the intuition that it is permissible for me to take the left path in the original case.

So, I think that Rossian Minimalism is subject to the Walk in the Park Objection just as much as World Minimalism is. After all, both theories maintain that the permissibility of an action depends solely on the non-mental facts of the agent's situation. But any such theory will be subject to objections like the Walk in the Park Objection. For any such theory will make it possible for the facts to obtain that make an action permissible even though its agent doesn't believe those facts obtain, or has no reason to believe they obtain. And to some, this will result in unintuitive implications.

We've seen, then, that while Rossian Minimalism can solve some of the problems that affect other Rossian moral theories, it can't solve them all. Moreover, I'll argue in the next section that Rossian Minimalism has several problems that other Rossian moral theories, in particular World Rossianism, do not have.

5.7. Other Problems for Rossian Minimalism

It seems that there can be cases where a person has more prima facie duty *to* do one thing rather than another but no prima facie duty *not* to do either. In a case like this, Rossian Minimalism implies that each option is permissible. However, this can sometimes seem counterintuitive.

Consider, for instance, the following case. Suppose I can either perform A or perform B. I have no prima facie duty not to perform either action: neither would bring about any badness, breaks any promises, etc. However, A would bring about 1,000 units of goodness (say, 1,000 units of welfare for others), whereas would B bring about no

goodness. If Rossian Minimalism is true, then each option is permissible since, by stipulation, I don't have any prima facie duty not to do either; each therefore minimizes prima facie wrongness. But this conclusion seems incorrect: given the choice between bringing about no goodness and bringing about a lot of goodness (for others), it intuitively seems that I should do the latter (assuming, of course, that other things are equal).

Other examples of this kind of case can be given. Suppose, for instance, that I can either give \$1,000 to you, or give it to a friend. I have no prima facie duty not to perform either action. However, I have previously wronged you (I've crashed into your car, inflicting \$1,000 of damage to it), but I haven't previously wronged my friend. If Rossian Minimalism is true, then each of my options is permissible since each minimizes prima facie wrongness. But intuitively, I ought to give the money to you, not my friend (assuming, again, that everything else about my options is equal).

Or suppose you've done me a favor (say, you wrote me a nice letter of recommendation), and I can either write you a thank-you note, or go out of my way to thank you in person. The latter option, I'll assume, is an act of greater gratitude than the former. Neither option, however, is an act of *ingratitude*; each is an act of considerable gratitude. Suppose I have no prima facie duty not to perform either option. Rossian Minimalism then implies that each option is permissible. However, one might think that if everything else about my options is equal, I ought to perform the act of greater gratitude.

Notice that World Rossianism makes the intuitively correct conclusions about these scenarios. It's presumably the case that we have a prima facie duty to bring about

goodness. So, in the first example, I have a prima facie duty to perform A in virtue of this fact, but no such duty to perform B. If I have no prima facie duty not to perform either of my options, and everything else is equal, it could very well turn out that I perform A in each world accessible to me that maximizes net prima facie rightness. I would then be obliged to perform A, just as it seems. A similar line of reasoning shows that I could very well be obliged to give the \$1,000 to you in the second case and thank you in person in the third case, at least if we assume, as seems plausible, that there is a prima facie duty of compensation and a prima facie duty of gratitude (whose strength is proportional to the degree of gratitude).

There is, however, a way for a Rossian Minimalist to respond to these cases. She could say that the cases I have described are not possible. In the first case, I stipulated that I have no prima facie duty not to perform either A or B. However, I also said that A would bring about lots of goodness, whereas B would not. But a Rossian Minimalist could say that a person has a prima facie duty not to perform an action that fails to maximize goodness (where an action maximizes goodness iff it would bring about at least as much goodness as each of its alternatives would bring about). So, if A would bring about more goodness than B, then B does not maximize goodness, and I have a prima facie duty not to perform it in virtue of this fact. It therefore turns out that I have a prima facie duty not to perform B after all.

Similar things apply to the other cases I described. A Rossian Minimalist can say that a person has a prima facie duty not to fail to maximize compensation for her past wrongs (where an action maximizes compensation for a past wrong iff the compensation it provides for that wrong is at least as great as that of each of its alternatives). Likewise,

a person has a prima facie duty not to fail to maximize gratitude (where an action maximizes gratitude iff the gratitude it provides is at least as great as that of each of its alternatives). Given this, I *do* have a prima facie duty not to give the \$1,000 to my friend in the second case (since doing this fails to maximize compensation for a past wrong), and I *do* have a prima facie duty not to send you a thank you card in the third (since doing this fails to maximize gratitude). But this contradicts my assumption that I have no prima facie duty not to perform either of my options in these cases.

I have a few things to say in reply. First, the prima facie duties the Rossian Minimalist appeals to here are rather unusual. It's difficult to think of an independent reason for thinking that they exist. There's a worry, then, that the Minimalist's response is *ad hoc*.

Second, a World Rossian need not posit the existence of these unusual duties. She can make do with the much more natural prima facie duties of beneficence, compensation, and gratitude. That World Rossianism allows a Rossian to accept a less exotic package of prima facie duties is, I think, a consideration in its favor.

Third, if a Rossian Minimalist appeals to these unusual prima facie duties, then I think she'll face a version of the Procrastination Problem. Let's return to the case involving compensation. Suppose I can either pay you the \$1,000 right now, or eat lunch instead. (If I eat lunch now, I'll pay you after lunch.) Your car will be at the shop for a few days; it makes no difference to you whether I pay you now or after lunch. But notice that if I eat lunch now, I fail to maximize compensation for a past wrong. Eating lunch now provides no compensation to you, whereas an alternative to it (paying you the \$1,000 now) does. Eating lunch now therefore does not compensate you at least as much

as that of each alternative. So, if a Rossian Minimalist accepts that we have a prima facie duty not to fail to maximize compensation, it could very well turn out that eating lunch now fails to minimize prima facie wrongness. She'll then be forced to say that I ought to pay you now instead of eating lunch. But that doesn't seem correct.

Similar things apply to the duty of not failing to maximize gratitude. Suppose you did a great favor for me yesterday. I could either thank you for it today, or work today. (If I work today, I will thank you tomorrow.) It would make no difference to you which day I thank you; you'd be equally pleased either way. If a Rossian Minimalist accepts that we have a prima facie duty not to fail to maximize gratitude, then she'll have to admit that I have a prima facie duty not to work today. It could then turn out that working today fails to minimize prima facie wrongness, and a Rossian Minimalist will have to say, implausibly, that I'm obliged to thank you today instead of working today.

Here's the point. The original Procrastination Problem arose because it seems that we have a prima facie duty to keep our promises, but promise keepings are procrastinable. Since a Rossian Minimalist denies that permissibility is determined by our prima duties *to* do things, she avoids the Procrastination Problem. A similar problem does not arise when we consider promise breaking, for while it seems that we have a prima facie duty not to break a promise, promise breakings *aren't* procrastinable. We are never allowed to "put off" breaking a promise, as it were. Indeed, this doesn't even make sense: if you can "put off" breaking a promise, then you aren't breaking the promise after all. Rossian Minimalism is therefore an elegant solution to the problem. However, in order to yield the intuitively correct conclusions about a wide variety of cases, a Rossian Minimalist is going to have to accept a wide variety of prima facie duties not to do things.

And, as I've shown, some of these duties are bound to be procrastinable. By accepting them, she will thus face a version of the Procrastination Problem herself.

5.8. Conclusion

In this chapter, I have been challenging Markosian's claim that Rossian Minimalism is the best ethical theory that makes use of Ross's concept of a prima facie duty. I have been trying to show that World Rossianism—a "world" version of the moral theory Ross accepts in *R&G*—is more plausible. According to Markosian, World Rossianism faces some difficulties, in particular the Gratuitous Duty Objection and the Walk in the Park Objection. However, neither of these objections is very strong, and I've explained how a World Rossian should respond to them. In fact, Rossian Minimalism is also subject to at least one of these objections: a version of the Walk in the Park Objection can also be raised against it. Plus, Rossian Minimalism has other problems. It must contend with the Bergström/Castañeda Problems. And in order for a Minimalist to make the intuitively correct conclusions about a wide variety of cases, she'll have to combine Rossian Minimalism with some rather exotic prima facie duties not to do things, duties that appear procrastinable. By accepting them, a Rossian Minimalist will therefore encounter the Procrastination Problem. On the whole, then, I find World Rossian superior to Rossian Minimalism.

CHAPTER 6

THE PROBLEM OF SELF-BENEFIT AND SELF-HARM

6.1. Introduction

In the previous chapter, we saw that the moral theory that Ross accepts in *R&G* (namely, RTO_{RG} - RTW_{RG}) is subject to the Bergström/Castañeda Problem. To solve this problem, I argued that proponents of RTO_{RG} - RTW_{RG} should revise their theory to World Rossianism. In this chapter, I will consider another important objection to RTO_{RG} - RTW_{RG} , one that's inspired by Michael Stocker (1976) and Michael Slote (1984).

To see the objection I have in mind, consider the following case:

Jolt of Pleasure

Suppose I'm sitting at a table with a green button and a white button in front of me. I have only two options: I can press the green button, or I can press the white one. If I press the green button, I will receive a jolt of mild pleasure (perhaps equivalent to the amount of pleasure I receive when I take a hot shower for a few minutes). Let's suppose (as Ross does in *R&G*) that I have a prima facie duty to benefit myself (i.e., a prima facie duty of self-improvement). If that's the case, then I have a prima facie duty to press the green button. Suppose that this is the only prima facie duty I have to (or not to) perform this option. So, the net prima facie rightness of pressing the green button is, let's say, +5. On the other hand, if I press the white button, nothing much will happen. Let's suppose that I therefore have no prima facie duty to (or not to)

perform this option. The net prima facie rightness of pressing the white button is therefore 0.

According to RTO_{RG} , I am morally obligated to press the green button and thereby give myself a jolt of pleasure since doing so uniquely maximizes net prima facie rightness. However, I suspect that many will find this implication counterintuitive. I suspect that many will want to say that while I am *permitted* to press the green button, I am not *obligated* to do so. Perhaps it would, in some sense, be silly for me not to press the green button in this case. Maybe it would even be irrational for me not to do so. But surely it wouldn't be morally wrong.

Here's a related objection. Consider this case:

Jolt of Pain

Suppose I'm again sitting at a table, but this time I have a red button and a white button in front of me. I have only two options: I can press the red button, or I can press the white one. If I press the red button, I will receive a jolt of mild pain (perhaps equivalent to the amount of pain I receive when I get a flu shot). Let's suppose that I have a prima facie duty not to harm myself (i.e., a prima facie duty against self-destruction). If that's the case, then I have a prima facie duty not to press the red button. Suppose that this is the only prima facie duty I have not to (or to) perform this option. So, its net prima facie rightness is, we'll suppose, -5. On the other hand, if I press the white button, nothing much will happen, and so its net prima facie rightness is 0.

According to RTW_{RG} , it's morally wrong for me to press the red button and thereby give myself a jolt of pain since doing so fails to maximize net prima facie rightness. However, I suspect that many will find this difficult to accept. I suspect that many will claim that while it would be bizarre, perhaps even irrational, for me to press the red button in this case, it wouldn't be morally wrong for me to do so.

In this chapter, I will explore various ways of responding to these objections. I'll argue that a proponent of RTO_{RG} - RTW_{RG} can handle them, though to do so, she'll need to make another modification to her theory.

6.2. First Response: Embrace the Implications

One way for a proponent of RTO_{RG} - RTW_{RG} to respond to these objections is to simply insist that I am morally obligated to press the green button in Jolt of Pleasure and that it's morally wrong for me to press the red button in Jolt of Pain. This seems to be how Ross, in his early work, would have responded to these objections. However, I suspect that most proponents of RTO_{RG} - RTW_{RG} will be uncomfortable with this line of response. Most proponents will, I suspect, take inspiration from the following memorable passages from *The Right and the Good*:

The main moral convictions of the plain man seem to me to be, not opinions which it is for philosophy to prove or disprove, but knowledge from the start. (1930, 20-21n)

The moral convictions of thoughtful and well-educated people are the data of ethics just as sense-perceptions are the data of natural science. (41)

Loyalty to the facts is worth more than a symmetrical architectonic or a hastily reached simplicity. (23)¹

I therefore suspect that most proponents of RTO_{RG} - RTW_{RG} —and, for that matter, most theorists in the broadly Rossian tradition—will want their moral theory to cohere with commonsense morality as much as possible: that seems to be the overarching Rossian methodology. However, it seems to be part of the common moral consciousness that I am not obliged to press the green button in Jolt of Pleasure and it's not wrong for me to press the red button in Jolt of Pain. So, I suspect that most proponents of RTO_{RG} - RTW_{RG} will want to accommodate these intuitions instead of flatly rejecting them, even if this is what the early Ross would have done.

6.3. Second Response: Appeal to Praiseworthiness and Blameworthiness

Another way for a proponent of RTO_{RG} - RTW_{RG} to respond to the objections in question is to deny that it's part of commonsense morality that I am not obliged to press the green button in Jolt of Pleasure and that it's not wrong for me to press the red button in Jolt of Pain. A proponent of RTO_{RG} - RTW_{RG} might argue that what is part of commonsense morality is not that I'm not obliged to press the green button in Jolt of Pleasure, but rather, that I wouldn't be *praiseworthy* for pressing it; if someone thinks that I'm not obliged to press the button, that's only because one is confusing obligation with praiseworthiness. Similarly, a proponent of RTO_{RG} - RTW_{RG} might argue that what is part of commonsense morality is not that it's not wrong for me to press the red button in Jolt of Pain, but rather that I wouldn't be *blameworthy* for pressing it; if someone

¹ Ross says similar things in *FE* (see, esp., 1939, 1, 3).

thinks that it's not wrong for me to press the button, that's only because one is confusing wrongdoing with blameworthiness.

However, I again doubt that many proponents of RTO_{RG} - RTW_{RG} will want to make this response. Consider again the case Ross uses to object to Ideal Utilitarianism:

Promise Keeping

Suppose I can either do A or do B. Performing A will bring about 1000 units of net intrinsic goodness. On the other hand, performing B will bring about 1001 units of net intrinsic goodness. However, if I do A, I will keep a promise, but if I do B, I will break a promise.

According to Ideal Utilitarianism, I ought to do B since doing it uniquely maximizes net intrinsic goodness. However, according to Ross, commonsense morality disagrees with Ideal Utilitarianism here. According to the common moral consciousness, I should do A: keeping a promise is "worth more," morally speaking, than one additional unit of goodness.²

But notice that an Ideal Utilitarian can respond to this difficulty in the same way that we are currently responding to the difficulties presented by Jolt of Pleasure/Pain. That is, an Ideal Utilitarian can say that what is part of the commonsense moral consciousness is *not* that it would be wrong for me to break my promise in Promise Keeping, but rather, that I would be blameworthy for doing so; if someone thinks it's wrong for me to break my promise, that's only because one is confusing wrongdoing with blameworthiness.

² I discuss this objection in sec. 1.2.

This is a common way for a utilitarian to respond to problem cases like Promise Keeping. Moore, for instance, appeals to it in at least one instance ([1912] 2005, 97-101). However, I suspect that proponents of RTO_{RG} - RTW_{RG} —and Rossian theorists in general—will not be impressed by it (Ross evidently wasn't). Indeed, they will likely say that cases like Promise Keeping are precisely the kinds of cases that push us to abandon utilitarianism and accept a Rossian moral theory instead. However, if a proponent of RTO_{RG} - RTW_{RG} holds that utilitarians can't successfully respond to Promise Keeping by appealing to the distinction between obligation/wrongdoing and praiseworthiness/blameworthiness, then it's difficult to see what principled grounds she could have for thinking that the distinction can help her respond to Jolt of Pleasure/Pain.

6.4. Third Response: Deny the Prima Facie Duties

Notice that in Jolt of Pleasure and Jolt of Pain, we assumed, respectively, that there's a prima facie duty to benefit oneself and a prima facie duty not to harm oneself. However, if a proponent of RTO_{RG} - RTW_{RG} denies this assumption (and Ross does just that in *FE*), then these cases won't trouble her. If there is no duty to (not to) benefit (harm) oneself, then RTO_{RG} - RTW_{RG} implies that it's permissible but not obligatory for me to press the green button in Jolt of Pleasure, and it's permissible (i.e., not wrong) for me to press the red button in Jolt of Pain, just as commonsense morality indicates.

Unfortunately, if a proponent of RTO_{RG} - RTW_{RG} responds to the Jolt of Pleasure and Jolt of Pain objections in this way, she'll face a different set of problems. Consider this case:

Money

I find myself in an unusual situation where I have two nice options: I can either ensure that *you* get \$5, or I can ensure that *I* get \$10,000 (that I can use only on myself). I presumably have a prima facie duty to secure the \$5 for you since there is a prima facie duty to benefit others. On the other hand, since we're currently assuming that there is no prima facie duty to benefit oneself, the fact that securing the \$10,000 for me would benefit me greatly does not give me any prima facie duty to do it. Let's suppose that there are no other prima facie duties involved in my scenario. If that's the case, then the net prima facie rightness of ensuring that you get the \$5 will be positive, say +5, and the net prima facie rightness of ensuring that I get the \$10,000 will be 0.

$R_{TO_{RG}} - R_{TW_{RG}}$ implies that I am obligated to secure the \$5 for you; it's wrong for me to secure the \$10,000 for myself. But this is difficult to accept. Surely I'm at least permitted to secure the \$10,000 for myself. I may also be permitted to secure the \$5 for you instead, but intuitively, I'm not obligated to do so.

Here's another, related case:

Appendages

I again find myself in unusual situation, this time with two grisly options: I can either cut off your right index finger, or I can chop off my right arm. I presumably have a prima facie duty not to cut off your finger since there is a prima facie duty not to harm others. However, since we're assuming that there is no prima facie duty not to harm oneself, the fact that chopping off my arm will harm me greatly does not give me any prima facie duty not to do it. Suppose that there are no other prima facie duties involved in the case. The net prima facie rightness of cutting off your finger

will thus be negative, say -100, while the net prima facie rightness of chopping off my arm will be 0.

According to RTO_{RG} - RTW_{RG} , I am obligated to chop off my arm; it's wrong for me to chop off your finger instead. But surely this is mistaken. I may be permitted to chop off my own arm here, but surely I'm not obligated to do so.

It seems, then, that what a proponent of RTO_{RG} - RTW_{RG} needs to do is to deny that there is a prima facie duty to (not to) benefit (harm) oneself while at the same time maintaining that self-benefit and self-harm are relevant to determining the deontic status of an action. How would she go about doing this? In the next section, I'll suggest a way, a way that's inspired by some of Doug Portmore's work (especially his 2008, 2011).

6.5. Fourth Response: Dual-Ranking

Recall what a prima facie duty is. To say that there is a prima facie duty for a person to do A is to say that there is a moral reason for the person to do A—that is, there is a moral consideration that counts in favor of the person's doing A (see sec. 1.3). Given that a prima facie duty just is a *moral* reason, we could also call a prima facie duty a “prima facie *moral* duty.” Jolt of Pleasure and Jolt of Pain demonstrate, I think, that there is no prima facie moral duty to benefit oneself and no prima facie moral duty not to harm oneself. However, there is surely a prima facie *prudential* duty to (not to) benefit (harm) oneself. To say that there is a prima facie prudential duty for a person to do A is to say that there is a prudential reason for the person to do A—that is, there is a prudential consideration that counts in favor of the person's doing A. Prudential reasons have to do with a person's welfare or well-being, and the fact that an action would benefit

(harm) the agent is a paradigmatic example of a prudential reason for the agent to (not to) do it. And while Jolt of Pleasure and Jolt of Pain demonstrate that there is no prima facie moral duty to benefit oneself and no prima facie moral duty not to harm oneself, Money and Appendages demonstrate, I think, that the moral status of an action is a function of not only one's prima facie moral duties, but also one's prima facie prudential duties.

What function, in particular? As a first step toward answering this question, consider the following theory:

Dual-Ranking Rossianism³

An action A is morally obligatory for S(A) to perform if and only if (and because) A uniquely maximizes both the net balance of prima facie moral duties and the net balance of prima facie prudential duties;

An action A is morally permissible for S(A) to perform if and only if (and because) A maximizes either the net balance of prima facie moral duties or the net balance of prima facie prudential duties;

An action A is morally wrong for S(A) to perform if and only if (and because) S(A) can perform A, and A maximizes neither the net balance of prima facie moral duties, nor the net balance of prima facie prudential duties;

where

An action A maximizes the net balance of prima facie moral duties =_{df.} S(A) can perform A, and [the sum of the strengths of all of the prima facie moral duties S(A) has to perform A] – [the sum of the strengths of all of the prima facie moral duties S(A) has not to perform A] is at least as great as that of any other action S(A) can perform instead;

An action A uniquely maximizes the net balance of prima facie moral duties =_{df.} S(A) can perform A, and [the sum of the strengths of all of the prima facie moral duties

³ I call this theory “Dual-Ranking Rossianism” because it is similar to a version of act consequentialism that Portmore (2008, 2011) calls “Dual-Ranking Act-Consequentialism.”

$S(A)$ has to perform A] – [the sum of the strengths of all of the prima facie moral duties $S(A)$ has not to perform A] is greater than that of any other action $S(A)$ can perform instead;

An action A maximizes the net balance of prima facie prudential duties =_{df.} [the *definiens* of “action A maximizes the net balance of prima facie moral duties” except that all instances of “moral” are replaced with “prudential”];

Action A uniquely maximizes the net balance of prima facie prudential duties =_{df.} [the *definiens* of “action A uniquely maximizes the net balance of prima facie moral duties” except that all instances of “moral” are replaced with “prudential”].

If a proponent of RTO_{RG} - RTW_{RG} modifies her theory to Dual-Ranking Rossianism, then she can say the intuitively correct things about each of the above cases. Let’s return briefly to each of these cases.

Jolt of Pleasure. In this case, I can press the green button and thereby give myself a jolt of pleasure, or I can press the white button and nothing much will happen. I’m now assuming that the fact that an action of mine will benefit me gives me no prima facie moral duty to do it; instead it gives me a prima facie prudential duty to do it. Since I’m also assuming that there are no other prima facie moral or prudential duties involved in my scenario, it follows that pressing the green button (uniquely) maximizes the net balance of prima facie prudential duties, but each of my options maximizes the net balance of prima facie moral duties. The following chart summarizes this information:

Table 4: Jolt of Pleasure

<u>Alternative</u>	<u>Net balance of prima facie moral duties</u>	<u>Net balance of prima facie prudential duties</u>	<u>Maximizes net balance of prima facie moral duties?</u>	<u>Maximizes net balance of prima facie prudential duties?</u>
Press green button	0	+5	Yes	Yes (uniquely)
Press white button	0	0	Yes	No

Since each of my options maximizes either the net balance of prima facie moral duties or the net balance of prima facie prudential duties, each option is permissible according to Dual-Ranking Rossianism. Moreover, since neither option uniquely maximizes both the net balance of prima facie moral duties and the net balance of prima facie prudential duties, neither is obligatory according to the theory.

Jolt of Pain. Here I can press the red button and thereby give myself a jolt of pain, or I can press the white button and nothing will happen. I'm assuming that the fact that an action of mine will harm me gives me no prima facie moral duty not to do it; however, it does give me a prima facie prudential duty not to do it. Since this, by stipulation, is the only moral or prudential duty involved in my scenario, pressing the white button (uniquely) maximizes the net balance of prima facie prudential duties, but each of my options maximizes the net balance of prima facie moral duties. This can be seen in the following chart:

Table 5: Jolt of Pain

<u>Alternative</u>	<u>Net balance of prima facie moral duties</u>	<u>Net balance of prima facie prudential duties</u>	<u>Maximizes net balance of prima facie moral duties?</u>	<u>Maximizes net balance of prima facie prudential duties?</u>
Press red button	0	-5	Yes	No
Press white button	0	0	Yes	Yes (uniquely)

According to Dual-Ranking Rossianism, each of my options is permissible since each maximizes either the net balance of prima facie moral duties or the net balance of prima facie prudential duties; however, neither is obligatory since neither uniquely maximizes both of these things.

Money. In this case, I have the option of securing \$5 for you or \$10,000 for me. By assumption, the fact that an action of mine would benefit me greatly gives me no prima facie moral duty to do it; however, it does give me a strong prima facie prudential duty to do it. On the other hand, the fact that an action of mine would benefit you a little presumably does give me a (relatively weak) prima facie moral duty to do it. Assuming that there are no other pertinent features of the scenario, the following chart summarizes the prima facie moral and prudential duties involved in it:

Table 6: Money

<u>Alternative</u>	<u>Net balance of prima facie moral duties</u>	<u>Net balance of prima facie prudential duties</u>	<u>Maximizes net balance of prima facie moral duties?</u>	<u>Maximizes net balance of prima facie prudential duties?</u>
Secure \$5 for you	+5	0	Yes (uniquely)	No
Secure \$10,000 for me	0	+10,000	No	Yes (uniquely)

As the chart indicates, each of my options is permissible according to Dual-Ranking Rossianism, but neither is obligatory.

Appendages. Here I have the option of cutting off your right finger or cutting off my right arm. I'm assuming that the fact that an action of mine would harm me severely gives me no prima facie moral duty not to do it; instead, it gives me a (very weighty) prima facie prudential duty not to do it. On the other hand, the fact that an action of mine would harm you to a lesser extent presumably does give me a (somewhat less weighty) prima facie moral duty not to do it. Given that there is nothing else of interest about the situation, here's a chart depicting the prima facie moral and prudential duties it involves:

Table 7: Appendages

<u>Alternative</u>	<u>Net balance of prima facie moral duties</u>	<u>Net balance of prima facie prudential duties</u>	<u>Maximizes net balance of prima facie moral duties?</u>	<u>Maximizes net balance of prima facie prudential duties?</u>
Cut off your right finger	-1,000	0	No	Yes (uniquely)
Cut off my right arm	0	-10,000	Yes (uniquely)	No

As can be seen from the chart, Dual-Ranking Rossianism implies that neither of my options is obligatory; each is permissible.

6.6. Dual-Ranking, Take 2

While Dual-Ranking Rossianism yields the intuitively correct implications about each of the cases discussed so far, it has a fatal problem. Consider this case:

World Poverty

Suppose I find myself in an astonishing scenario where I can either watch that season finale on TV that I've been dying to watch, or I can end world hunger. Watching the TV show will benefit me a little, and so, by assumption, I have a prima facie prudential duty to do it in virtue of this fact, but not a prima facie moral duty. On the other hand, ending world hunger will benefit others tremendously, and so I presumably have an enormously strong prima facie moral duty to do it. Assume that there are no other prima facie moral or prudential duties involved in the case (so, assume that ending world hunger will not benefit me in any way).

Here's a chart that summarizes this information:

Table 8: World Poverty

<u>Alternative</u>	<u>Net balance of prima facie moral duties</u>	<u>Net balance of prima facie prudential duties</u>	<u>Maximizes net balance of prima facie moral duties?</u>	<u>Maximizes net balance of prima facie prudential duties?</u>
Watch TV	0	+10	No	Yes (uniquely)
End world hunger	+1 zillion	0	Yes (uniquely)	No

As the chart indicates, Dual-Ranking Rossianism implies that each of my options is permissible since watching TV maximizes the net balance of prima facie prudential duties, and ending world hunger maximizes the net balance of prima facie moral duties. However, this implication is counterintuitive. Surely I am obligated to end world hunger in this case; it would be wrong for me to watch TV instead. Dual-Ranking Rossianism, then, needs modification.

Here's an initially plausible way of repairing the theory:

Dual-Ranking Rossianism2

An action A is morally obligatory for S(A) to perform if and only if (and because) A uniquely maximizes both the net balance of prima facie moral duties and the net balance of prima facie prudential duties + **the net balance of prima facie moral duties**;⁴

An action A is morally permissible for S(A) to perform if and only if (and because) A maximizes either the net balance of prima facie moral duties or the net balance of prima facie prudential duties + **the net balance of prima facie moral duties**;

⁴ Here I am obviously assuming that it makes sense to add together prima facie moral duties and prima facie prudential duties.

An action A is morally wrong for S(A) to perform if and only if (and because) S(A) can perform A, and A maximizes neither the net balance of prima facie moral duties, nor the net balance of prima facie prudential duties + **the net balance of prima facie moral duties**.

To see how Dual-Ranking Rossianism2 handles World Hunger, consider the following updated chart:

Table 9: World Hunger according to Dual-Ranking Rossianism2

<u>Alternative</u>	<u>Net balance of prima facie moral duties</u>	<u>Net balance of prima facie prudential duties</u>	<u>Net balance of prima facie prudential + prima facie moral duties</u>	<u>Maximizes net balance of prima facie moral duties?</u>	<u>Maximizes net balance of prima facie prudential + prima facie moral duties?</u>
Watch TV	0	+10	+10	No	No
End world hunger	+1 zillion	0	+1 zillion	Yes (uniquely)	Yes (uniquely)

As can be seen from the chart, watching TV fails to maximize the net balance of prima facie moral duties, and it fails to maximize the net balance of prima facie prudential + prima facie moral duties, so it's wrong according to Dual-Ranking Rossianism2. Ending world hunger, on the other hand, uniquely maximizes both the net balance of prima facie moral duties and the net balance of prima facie prudential + prima facie moral duties, so it's obligatory according to Dual-Ranking Rossianism2. Furthermore, Dual-Ranking Rossianism2 implies the intuitively correct things about Jolt of Pleasure, Jolt of Pain, Money, and Appendages, though I'll leave it to the reader to verify this for herself.

6.7. Dual-Ranking, Take 3

Unfortunately, Dual-Ranking Rossianism² is problematic. Consider this case:

Fire

A house is burning down. Two children are trapped inside. I'm the only one who can help. I have three choices: I can do nothing, I can run into the house and save one of the children, or I can run into the house and save both children. If I do nothing, both children will die. If I run into the house and save one child, I will sustain severe burns, which will result in my losing an arm. If I run into the house and save both children, I will suffer even worse burns, which will result in my losing both of my arms.

Let's make the following assumptions about the prima facie moral and prudential duties involved in the case:

1. *Do nothing.* I have a strong prima facie moral duty against performing this option in virtue of the fact that it will result in the death of two children. On the other hand, I have a strong prima facie prudential duty to perform this option in virtue of the fact that it will save both of my arms.
2. *Enter house, save only one child.* I have a strong prima facie moral duty to do this in virtue of the fact that it will save a child. However, I also have a strong prima facie moral duty against doing this in virtue of the fact that it will result in the death of a child. (I'll assume that these duties balance out.) On the other hand, I have a strong prima facie prudential duty to perform this option in virtue of the fact that it will save one of my arms. However, I also have a strong prima

facie prudential duty not to do it in virtue of the fact that it will result in my losing an arm. (I'll assume that these duties also balance out.)

3. *Enter house, save two children.* I have a very strong prima facie moral duty to do this in virtue of the fact that it will save two children. However, I also have a very strong prima facie prudential duty not to do this in virtue of the fact that it will result in my losing both of my arms.
4. There are no other prima facie moral or prudential duties involved in the case.

Here's a chart summarizing this information⁵:

⁵ These numbers in the following chart are not to scale with the numbers I assigned to the strengths of prima facie duties elsewhere in this chapter. For instance, I earlier said that the prudential duty against my losing one of my arms is -10,000, but now I'm saying that it's only -10. I've made the numbers smaller in the present case to prevent them from getting unwieldy.

Table 10: Fire

<u>Alternative</u>	<u>Net balance of prima facie moral duties</u>	<u>Net balance of prima facie prudential duties</u>	<u>Net balance of prima facie prudential + prima facie moral duties</u>	<u>Maximizes net balance of prima facie moral duties?</u>	<u>Maximizes net balance of prima facie prudential + prima facie moral duties?</u>
Do nothing	-200	+20	-180	No	No
Enter house, save one child	0 (100-100)	0 (10-10)	0	No	No
Enter house, save both children	+200	-20	+180	Yes (uniquely)	Yes (uniquely)

Notice, however, that in this case, saving both children uniquely maximizes both the net balance of prima facie moral duties and the net balance of prima facie prudential + prima facie moral duties. Dual-Ranking Rossianism² therefore implies that it is obligatory and my other options are wrong. But that is counterintuitive. Presumably it's permissible for me to do nothing in this case. It's also presumably permissible for me to save only one child; if I am permitted to make a huge personal sacrifice in order to save both children, then surely I'm also permitted to make a lesser personal sacrifice in order to save only one of them.

What causes the problem for Dual-Ranking Rossianism² in this case is that the strength of my prima facie moral duty not to sacrifice a child swamps my prima facie prudential duty not to sacrifice my arm: I assumed that the strength of the former duty is 100 but the strength of the latter is only 10. Naturally, then, one way for a Dual-Ranking

Rossian2 to respond to the problem is to argue that this assumption is false. She might say, in particular, that the strength of my prima facie prudential duty not to sacrifice my arm is significantly greater than 10. Suppose, for instance, that it's 500. Then the above chart would look like this:

Table 11: Fire, updated

<u>Alternative</u>	<u>Net balance of prima facie moral duties</u>	<u>Net balance of prima facie prudential duties</u>	<u>Net balance of prima facie prudential + prima facie moral duties</u>	<u>Maximizes net balance of prima facie moral duties?</u>	<u>Maximizes net balance of prima facie prudential + prima facie moral duties?</u>
Do nothing	-200	+1,000	+800	No	Yes
Enter house, save one child	0 (100-100)	0 (500-500)	0	No	No
Enter house, save both children	+200	-1,000	-800	Yes	No

Now Dual-Ranking Rossianism2 implies that it's permissible for me to enter the house and save both children, and it's permissible for me to do nothing. This is a better result.⁶

⁶ There's another way for a Dual-Ranker to reach this result. She might claim that the strengths of the prima facie moral and prudential duties are accurately calculated in the original chart; however, she might go on to say that permissible actions are ones that either maximize the net balance of prima facie moral duties or maximize the net balance of prima facie moral duties + [(the net balance of prima facie prudential duties) × (some factor F)]. If this factor is high enough, then the theory in question will also imply that it's permissible for me to do nothing, and it's permissible for me to save both children. Of course, if a Dual-Ranker adopts this line of response, we'll want to know precisely what F is, and I suspect that a Dual-Ranker will not be able to say much about it.

However, Dual-Ranking Rossianism² *still* implies that it's wrong for me to save only one child. But this, I've said, is counterintuitive. Fortunately, there is a way to solve this problem. Consider the following tweak to Dual-Ranking Rossianism²:

Dual-Ranking Rossianism³

An action A is morally obligatory for S(A) to perform if and only if (and because) A uniquely maximizes both the net balance of prima facie moral duties and the net balance of prima facie prudential + prima facie moral duties;

An action A is morally permissible for S(A) to perform if and only if (and because) **S(A) can perform A, and there is no alternative to A that has both a greater net balance of prima facie moral duties and a greater net balance of prima facie prudential + prima facie moral duties;**

An action A is morally wrong for S(A) to perform if and only if (and because) S(A) can perform A, and **there is an alternative to A that has both a greater net balance of prima facie duties and a greater net balance of prima facie prudential + prima facie moral duties.**

According to Dual-Ranking Rossianism³, entering the house and saving only one child is permissible since there is no alternative to it that has both a greater net balance of prima facie moral duties and a greater net balance of prima facie prudential + prima facie moral duties. There is, of course, an alternative that has a greater net balance of prima facie moral duties (namely, saving both children), and there is an alternative that has a greater net balance of prima facie prudential + prima facie moral duties (namely, doing nothing). But no alternative is greater in both categories.

For similar reasons, Dual-Ranking Rossianism³ also implies that I am permitted to do nothing in Fire, and I am permitted to save both children. Moreover, I believe that the theory makes the intuitively correct conclusions about the other cases discussed in this chapter, though I will spare the reader the details. For these reasons, I think that a

proponent of $\text{RTO}_{\text{RG}}\text{-RTW}_{\text{RG}}$ would be wise to adopt Dual-Ranking Rossianism³, at least if she's interested in making the intuitively correct conclusions about the cases discussed here. However, let me reiterate that the success of Dual-Ranking Rossianism³—in particular, its success in handling cases like Fire—depends on the crucial assumption made above about the relative strengths of our prima facie moral and prudential duties. In particular, it depends on the assumption that our prudential duties are, in general, weighted more heavily than our moral duties.⁷ If this assumption is granted, however, then I think Dual-Ranking Rossianism³ is an attractive theory.

As an added bonus, if a proponent of $\text{RTO}_{\text{RG}}\text{-RTW}_{\text{RG}}$ modifies her theory to Dual-Ranking Rossianism³, she will have the resources to give a nice account of supererogation. A supererogatory action is one that “goes above and beyond the call of duty.” For instance, suppose I’m a soldier during WWI. A grenade has been thrown into my trench, and it’s about to explode. I can either run away from the grenade and save my own life, or jump on the grenade and sacrifice my life for the lives of several of my fellow soldiers. Intuitively, jumping on the grenade is supererogatory: I am permitted to do it, and it would be great if I did it, but morality doesn’t require me to do it. Doing it goes above and beyond the call of duty.

Following Zimmerman (1996, 235), I will assume that supererogatory actions have two important features. First, supererogatory actions are permissible but not obligatory: morally speaking, you *can* do them, but you don’t *have* to. Second, while supererogatory actions are permissible (but not obligatory), they are somehow “morally superior” to or

⁷ Or it depends on boosting the strength of our prima facie prudential duties by some factor (see note 6).

“morally more valuable” than other permissible alternatives. There will undoubtedly be debate among theorists about what it takes for one permissible alternative to be “morally superior” or “morally more valuable” than another. However, a Dual-Ranking Rossian³ has an account readily available to her. She can say that one permissible alternative is morally superior to another when the former’s net balance of prima facie moral duties is greater than the latter’s. More precisely, she can accept the following:

DRR3_{Sup}: An action A is supererogatory for S(A) to perform if and only if (a) S(A) is morally permitted to perform A; (b) there is some other alternative, B, that S(A) is also morally permitted to perform; and (c) A’s net balance of prima facie moral duties is greater than B’s.

Let’s apply this to Fire. According to DRR3_{Sup}, running into the burning house and saving one child is supererogatory for me, as is running into the house and saving two children: each of these actions is permissible, and each has a net balance of prima facie moral duties (0 and +200, respectively) that is greater than the net balance of another permissible alternative (namely doing nothing, which has a net balance of prima facie moral duties of -200). In Money, DRR3_{Sup} implies that securing \$5 for you is supererogatory for me since it is permissible and its net balance of prima facie moral duties (+5) is greater than the net balance of another permissible alternative (namely securing \$10,000 for myself, which has a net balance of prima facie moral duties of 0). For similar reasons, DRR3_{Sup} implies that cutting off my arm (instead of cutting off your finger) is supererogatory for me in Appendages. All of these implications seem correct to me. That Dual-Ranking Rossianism³ provides one with the resources for an intuitively plausible account of supererogation is, I think, an additional reason for a proponent of RTO_{RG}-RTW_{RG} to embrace it.

6.8. Appendix: Combining Dual-Ranking Rossianism³ and World Rossianism

In this chapter, I have argued that in light of the problems that arise from Jolt of Pleasure, Jolt of Pain, and other similar cases, a proponent of $\text{RTO}_{\text{RG}}\text{-RTW}_{\text{RG}}$ should modify her theory to Dual-Ranking Rossianism³. In chapter 5, I argued that in light of the Bergström/Castañeda Problem, a proponent of $\text{RTO}_{\text{RG}}\text{-RTW}_{\text{RG}}$ should modify her theory to World Rossianism. Of course, Dual-Ranking Rossianism³ and World Rossianism are quite different theories. So, what I actually think is that a proponent of $\text{RTO}_{\text{RG}}\text{-RTW}_{\text{RG}}$ should modify her theory to something that combines together the key features of Dual-Ranking Rossianism³ and World Rossianism. What would such a theory look like? Here's a proposal:

Dual-Ranking World Rossianism

It is morally obligatory, as of a time t , for $S(A)$ to perform an action A at $t(A)$ if and only if (and because) $S(A)$ performs A at $t(A)$ in each world that uniquely maximizes (for her at t) both the net balance of prima facie moral duties and the net balance of prima facie prudential + prima facie moral duties;

It is morally permissible, as of a time t , for $S(A)$ to perform an action A at $t(A)$ if and only if (and because) $S(A)$ performs A at $t(A)$ in some world that has the following feature: there is no other world accessible to $S(A)$ at t that has both a greater net balance of prima facie moral duties and a greater net balance of prima facie prudential + prima facie moral duties;

It is morally wrong, as of a time t , for $S(A)$ to perform an action A at $t(A)$ if and only if (and because) $S(A)$ performs A at $t(A)$ in a world accessible to her at t , but she doesn't perform it in a world that has the following feature: there is no other world accessible to her at t that has both a greater net balance of prima facie moral duties and a greater net balance of prima facie prudential + prima facie moral duties. (More simply, what this says is that it is wrong for an agent to do something iff she can do it but she is not morally permitted to do it.)

I believe that Dual-Ranking World Rossianism can solve the Bergström/Castañeda Problem while accommodating our commonsense moral intuitions in all of the cases presented in this chapter, something that no other theory discussed so far can do.

CHAPTER 7

ROSS'S THEORY OF SUBJECTIVE RIGHTNESS

7.1. Introduction

As we saw in chapter 4, Ross draws a distinction between the objective sense of “right” and the subjective sense of “right.” According to Ross, an action is right in the objective sense when it maximizes the net balance of prima facie rightness over wrongness, and an action is right in the subjective sense when the agent *believes* it maximizes the net balance of prima facie rightness over wrongness. More precisely:

- R3:** An action A is morally right (in the objective sense) for S(A) to perform if and only if (and because) S(A) can perform A, and A's total prima facie rightness minus its total prima facie wrongness is greater than that of any other action S(A) can perform instead;
- R4:** An action A is morally right (in the subjective sense) for S(A) to perform if and only if (and because) S(A) can perform A, and S(A) believes that A's total prima facie rightness minus its total prima facie wrongness is greater than that of any other action she can perform instead.

I agree with Ross that we should draw a distinction between the objective and the subjective senses of “right.” As we saw in sec. 4.2, Ross believes that we should draw this distinction because he thinks that the term “right” means “most morally suitable to the situation,” but there are different elements of an agent's situation, namely the objective element consisting of the non-mental facts of the agent's situation, and the subjective element consisting of the agent's mental state.

However, the distinction can be motivated in other ways as well. Consider, for instance, the following case:

Advil or Tylenol

Suppose a small child has a fever. The child's mother is trying to decide whether to give him some Advil or some Tylenol. After carefully reading the instructions on each box, the mother chooses, somewhat arbitrarily, to give her child the Tylenol.

However, unbeknownst to the mother, the Tylenol is defective. Tragically, the child has a severe allergic reaction to the defective Tylenol and dies. Had the mother given her child Advil instead, he would have been fine.¹

Question: has the mother acted wrongly in this case? Our intuitions pull us in different directions in cases like this one. (Or, at least, *mine* do.) On the one hand, the mother has killed her child, albeit accidentally. Upon realizing that the medicine she administered to her child resulted in his death, she will surely feel profound regret and may, in unbearable agony, think to herself, "I never should have given my child that Tylenol. I should have given him Advil instead." This thought may seem true, even if the "should" here has a moral sense. It may therefore seem that the mother has indeed done something wrong. On the other hand, there was no way the mother could have predicted the tragedy that her action would cause. Upon seeing the mother after the tragedy, one might be inclined to say something like this to her: "What happened was terrible, and I can understand why you are so distraught, but you should know that given your beliefs about your situation,

¹ For similar cases, see Zimmerman (2006; 2008, chap. 1) and Graham (2010, 93).

you did nothing wrong.” This may also seem true, even if the “wrong” here has a moral sense. It may therefore seem that the mother has not in fact acted wrongly.

So, our intuitions simultaneously push us to say that the mother has and hasn’t acted wrongly. And one natural way of responding to this tension is to posit the existence of *two* senses of “wrong” (and two corresponding senses of “right”), an objective sense and a subjective sense. Then we can say that the mother has in fact acted wrongly (where “wrong” here expresses the term’s objective sense), but we can also say that she’s done nothing wrong (where “wrong” here expresses the term’s subjective sense).

While this isn’t the only way to accommodate the intuitive tension that this case generates, and not everyone is willing to draw a distinction between the objective and the subjective senses of “right” (or “wrong”), for the remainder of this chapter I will simply take the distinction for granted.² My goal here is to explore the nature of subjective rightness,³ assuming that there is such a thing. I will begin by looking at Ross’s theory and argue that it is problematic. Then I will suggest a better theory in its place.

According to the theory I propose, an action is subjectively right iff it minimizes expected objective wrongness. After explaining this theory in detail, I will discuss some

² Like Ross, many contemporary philosophers draw a distinction between the objective and the subjective sense of “right” (see, e.g., Brandt 1959, 360-67; Russell 1966; Parfit 1984, 25; 2011, chap. 7; Feldman 1986; 2012; Gibbard 2005, 340; H. Smith 2010; Portmore 2011; Dorsey 2012). Indeed, this seems to be the dominant position these days. However, it’s certainly not uncontested (see, e.g., Thomson 1986; Zimmerman 1996; 2006; 2008; Graham 2010; Howard-Snyder 2005; Mason 2013).

³ What I really should say here is that I will explore the nature of the subjective sense of “right.” However, I will henceforth use the phrase “subjective sense of ‘right’” interchangeably with the phrase “subjective rightness.” I realize that by doing this, I am conflating senses and kinds, which, as Gary Matthews (1972) argues, we should be careful not to do. But in the present case, conflating these things increases ease of exposition, so I hope that Matthews will forgive my sins.

other theories of subjective rightness that have been suggested recently in the literature—including, in particular, Holly Smith's and Fred Feldman's—and argue that my theory is preferable. I will end by responding to some objections.

7.2. Some Alleged Problems for Ross's Theory, and Some Real Ones

I will begin by briefly discussing three problems for Ross's theory of subjective rightness (namely, R4) suggested by Zimmerman (2008, 13-16; 2006, 333-37) and McConnell (1988):

1. R4 implies that agents possess a certain sort of moral infallibility. If R4 is true and an agent knows it, and if agents always know what it is that they believe, then an agent can always know what it is that she is subjectively permitted to do (assuming that she doesn't make an inferential mistake). But it's absurd to hold that a person can determine what she can permissibly do simply by way of introspection.
2. If R4 is true, then it may well be the case that Hitler's actions during the Holocaust were subjectively right, since he may very well have believed that they maximize the net balance of prima facie rightness over wrongness. But that implication is difficult to accept.
3. The truth of R4 implies that if a person happens to have no beliefs about the net prima facie rightness of her actions, then none of her actions are subjectively right. But it is absurd to hold that merely failing to attend to one's situation (and thus failing to have any beliefs about the net prima facie rightness of one's actions) can leave one without any permissible options.

However, I doubt that these objections will sway many proponents of R4.⁴ With regard to the first of Zimmerman's and McConnell's complaints, proponents of R4 will likely say that it is no problem at all if it turns out that agents can always know which of their options are subjectively right. After all, many philosophers claim that action-guidance is a criterion of adequacy for any theory of subjective rightness (see, e.g., H. Smith 2010; Feldman 2012; Hedden 2012).⁵ So, proponents of R4 will likely say that Zimmerman's and McConnell's first problem with R4 is, in fact, a benefit of R4, not a problem at all.

With regard to the second of Zimmerman's and McConnell's complaints, proponents of R4 will likely admit that if Hitler really did believe that his actions during the Holocaust maximized total prima facie rightness over wrongness, then he may well have acted subjectively rightly. However, they'll say that this isn't all that counterintuitive, for we can still maintain that Hitler's actions are *objectively* wrong.

Proponents of R4 will likely respond to Zimmerman's and McConnell's third complaint in a similar fashion. They'll say that if a person really has no beliefs about the net prima facie rightness of her actions, then it may well be the case that none of her options are subjectively right. However, this isn't worrisome, for we can still hold that some of her options are *objectively* right.

⁴ To be fair to Zimmerman and McConnell, their objections are directed at Ross's theory of *obligation* from *FE*, not at his theory of subjective rightness. Directed at that target, their objections may be more convincing.

⁵ Later on, however, I will challenge this claim (see sec. 7.5).

For these reasons, I do not find the objections suggested by Zimmerman and McConnell all that convincing. However, I do think that R4 is problematic. Its primary problem, as I see it, is that it founders on Frank Jackson's well-known Dr. Jill case. Here's the case (suitably modified for the present context):

Dr. Jill

Dr. Jill has a patient, John, who is suffering from a minor skin ailment. Jill can give him one of three drugs: A, B, or C. She is certain that drug A will cure him partially. She is also sure that one of B or C will cure him completely and the other will kill him, but she doesn't know which is which. Jill is a committed objective Rossian; so, she believes that she should, objectively speaking, maximize the net balance of prima facie rightness over wrongness. In the present case, let's suppose, she knows that giving John drug A will *not* maximize the net balance of prima facie rightness over wrongness. Jill therefore thinks that giving A is objectively wrong. Still, she decides to give John drug A anyway because she wants to avoid the risk of killing him.⁶

Is Jill's action subjectively right? According to R4, it is not since Jill does not believe that giving drug A maximizes net prima facie rightness (indeed, she believes that it *fails* to maximize it). However, this is counterintuitive. While it may be *objectively* wrong for Jill to give drug A, surely it is *subjectively* right for her to do so.

⁶ See Jackson (1991, 462-63). Parfit (2011, 159-60) gives a similar case. An early version of the case can be found in Regan (1980, 264-65). Note that in the original version of the case, Jill is an objective utilitarian, not an objective Rossian.

A natural move to make at this point is to abandon R4 and replace it with one that makes use of something like expected utility.⁷ As we've seen (see sec. 2.5), the expected utility of an alternative is the sum, for each of its possible outcomes, of the outcome's actual value times the probability that it (the outcome) will obtain if the alternative is performed. More precisely, where A is an alternative, O_1, O_2, \dots, O_n are the possible outcomes of A, $V(O)$ is the value of outcome O, and $\text{prob}(O/A)$ is the probability of O given A, the expected utility of A = $\{[\text{prob}(O_1/A) \times V(O_1)] + [\text{prob}(O_2/A) \times V(O_2)] + \dots + [\text{prob}(O_n/A) \times V(O_n)]\}$. We can understand what might be called "expected net prima facie rightness" in a similar way: the expected net prima facie rightness of an alternative is the sum, for each possible balance of net prima facie rightness, of that balance times the probability that it (the alternative) has that balance. More precisely, where A is an alternative and B_1, B_2, \dots, B_n are the possible balances of net prima facie rightness of A, the expected net prima facie rightness of A = $\{[\text{prob}(A\text{'s net balance of prima facie rightness is } B_1) \times B_1] + [\text{prob}(A\text{'s net balance of prima facie rightness is } B_2) \times B_2] + \dots + [\text{prob}(A\text{'s net balance of prima facie rightness is } B_n) \times B_n]\}$.

There are different ways of understanding the sort of probability involved in expected utility/expected net prima facie rightness: we might understand it as being subjective probability, epistemic probability, or objective probability. Subjective probability has to do with credences or degrees of belief. So, the subjective probability that, say, I will win the lottery is the degree to which I believe that I will win. Epistemic probability has to do with the degree to which one's evidence supports something, or the

⁷ Indeed, this is how Jackson (1991) responds to the case.

degree of belief that one is justified in having. So, the epistemic probability that I will win the lottery is the degree to which I am justified in believing that I will win. Objective probability has to do with chances that exist as objective features of the world—chances that exist “in nature,” as it were. So, the objective probability that I will win the lottery is the chance that I will win, regardless of what I think that chance is or what my evidence suggests it is.⁸

I will assume that the type of probability involved in expected utility/expected net prima facie rightness is *epistemic* probability. In general (and contrary to Ross), it seems to me that the subjective moral status of an action is determined by the agent’s evidence, not her actual beliefs. However, not much turns on this. If one prefers to think that it is subjective probability (or objective probability) that is relevant to expected utility/expected net prima facie rightness, most of what I will say in this chapter should still be of interest: one can, without much loss, simply substitute talk of epistemic probability in what follows with talk of subjective (or objective) probability.

Let me now say a bit more about epistemic probability. I shall assume that for any proposition and for any person, we can assign a precise degree—between 0 and 1—to which the person is justified in believing the proposition. If a person is fully justified in believing a proposition, then we can say that the epistemic probability of the proposition

⁸ For helpful resources on the nature of probability, see Hájek (2012) and Mellor (2005). Note that Ross appears to think that there is no such thing as objective probability. He says, “But there cannot be probabilities in nature. Whatever the precise nature of the fact expressed by the statement ‘X has probably fainted’ may be, the fact must consist in our mind’s being in a certain state. Once this is realized it becomes clear that most of our ordinary thought involves the subjective view” (1939, 152). The last sentence of this passage is puzzling. Even if Ross is right about the non-existence of objective probability, it’s not clear to me how this would show that the subjective view is true.

for the person is 1. On the other hand, if a person is fully justified in believing that a proposition is false, then we can say that the epistemic probability of the proposition for the person is 0. If a person is just as justified in believing that a proposition is true as she is justified in believing that it is false, then we can say that the epistemic probability of the proposition for the person is .5. Of course, determining the precise conditions in virtue of which a person is justified in believing a proposition is an important philosophical task. Unfortunately, it's one that I will not be able to address here. I do want to note, however, that *being justified in believing* a proposition is crucially different from *justifiably believing* the proposition. The latter implies that one in fact believes the proposition in question. The former does not. If one is justified in believing a proposition, this merely implies, roughly, that *if one were* to believe the proposition, one *would* believe it justifiably.⁹ It does not imply, however, that one in fact believes it.

Consider, now, the following theory:

R4e: An action A is morally right (in the subjective sense) for S(A) to perform if and only if (and because) A maximizes expected net prima facie rightness,

where

⁹ I say that this is “roughly” correct because there might be cases where someone fails to believe p, has excellent evidence for believing p, but if she were to believe it, she would no longer have good evidence for it. What would such a case look like? Perhaps like this: suppose there's a cosmic evil demon who doesn't want anyone to believe in God. Mary doesn't believe in God (the evil demon rejoices) but she has good evidence for God's existence and is thus justified in believing in God. However, if she were to believe in God, the evil demon would have realized that Mary was about to believe in God, and he would have done all he could to prevent this from happening. In particular, he would have destroyed her good evidence for God's existence. So, although Mary is justified in believing in God, if she were to believe in God, she wouldn't believe in God justifiably.

An action A maximizes expected net prima facie rightness =_{df.} S(A) can perform A, and A's expected net prima facie rightness is at least as great as that of any other action S(A) can perform instead.

R4e yields the intuitively correct conclusions about the Dr. Jill case, at least after we make some plausible assumptions about it. Recall that Jill has three options: give drug A, give drug B, or give drug C. If she gives drug A, she knows that she will partially cure her patient, John. So, let's assume that Jill knows this option has a fairly high balance of net prima facie rightness (say, 90). For the purposes of illustration, let's assume that Jill's epistemic probability that giving A has this balance is 1. Given that, the expected net prima facie rightness of giving A is $1 \times 90 = 90$.

Let's turn now to Jill's second option: give drug B. Jill knows that this option will either kill John or cure him completely. So, let's assume that Jill knows that this option has either a very low balance of net prima facie rightness (say, -500), or a balance that is higher than that of her first option (say, 100). Jill, however, has no idea whether giving B will kill or perfectly cure her patient. So, let's assume, again for the purposes of illustration, that Jill's epistemic probability that giving B has the former balance is .5, and her epistemic probability that it has the latter is .5 as well. Given that, the expected net prima facie rightness of giving B is $(.5 \times 100) + (.5 \times -500) = -200$.

What I just said about Jill's second option also applies, *mutatis mutandis*, to Jill's third option: give drug C. I'll assume, then, that the expected net prima facie rightness of this option is also -200.

The following chart summarizes this information:

Table 12: Dr. Jill (expected utility)

<u>Alternative</u>	<u>Relevant possible balance of net prima facie rightness</u>	<u>Epistemic probability that this is the alternative's balance of net prima facie rightness</u>	<u>Expected net prima facie rightness of alternative</u>
Give drug A	90	1	$90 \times 1 = 90$
Give drug B	100 -500	.5 .5	$(100 \times .5) +$ $(-500 \times .5) =$ -200
Give drug C	100 -500	.5 .5	$(100 \times .5) +$ $(-500 \times .5) =$ -200

As indicated by the chart, giving drug A maximizes expected net prima facie rightness, but giving either of B or C does not. R4e therefore implies that the former option is subjectively right and the latter two are subjectively not right, just as it intuitively seems.

Unfortunately, although R4e nicely handles the Dr. Jill case, it has a different problem. Consider someone who accepts (and has good reason to accept) a non-Rossian theory of objective rightness, say Kantianism. According to a Kantian view of objective rightness, an action is objectively right iff its maxim is universalizable. Now suppose our person has two options: X and Y. He has overwhelming evidence that X (alone) has a universalizable maxim. Suppose, however, that Y (alone) maximizes expected net prima facie rightness. If R4e is true, then it is subjectively right for our person to do Y and subjectively wrong for him to do X. But this seems counterintuitive. Given that our person is justified in believing that X (alone) has a universalizable maxim, and given that

he has good reason to accept Kantianism about objective rightness, it seems subjectively right for him to do X and subjectively wrong for him to do Y.

The problem, in short, is that R4e is tailored too closely to Rossianism. It seems that a theory of subjective rightness should be relativized to whatever theory of objective rightness an agent is justified in accepting.¹⁰ R4e doesn't do that. To remedy this problem, one might be tempted to "disjunctivize" R4e in something like the following way:

R4ed: An action A is morally right (in the subjective sense) for S(A) to perform if and only if (and because) either

- (a) A maximizes expected net prima facie rightness and S(A) is justified in believing Rossianism about objective rightness, or
- (b) A maximizes expected universalizability and S(A) is justified in believing Kantianism about objective rightness, or
- (c) A maximizes expected utility and S(A) is justified in believing utilitarianism about objective rightness, or
- (d) A maximizes expected virtue exemplification and S(A) is justified in believing virtue ethics about objective rightness, or

...and so forth, for the rest of the theories of objective rightness that S(A) could be justified in accepting.

Of course, more needs to be said about R4ed before it can be fully understood. In particular, more needs to be said about the concepts of expected universalizability, expected virtue exemplification, etc. These concepts are intended to closely model the concepts of expected utility and expected net prima facie rightness, but it remains to be seen how exactly this modeling will work. I will not attempt to do that here, for even if it can be done, R4ed will remain an ugly monstrosity. I think many will rightly be

¹⁰ H. Smith (2010) and Feldman (2012) make a similar point.

suspicious of radically disjunctive theories like it. At the very least, it would be nice if the theory could be stated in a more succinct way. In the next section, I will propose such a way.

7.3. Minimizing Expected Objective Wrongness

It's part of commonsense morality that wrongness comes in degrees. In the vast majority of situations we face, we have many options, several of which are wrong, and it's plausible to hold that some of these are more wrong than others. Suppose, for instance, that I have three options: I can save one, I can kill one, or I can kill two. My latter two options are clearly wrong—at least *ceteris paribus*—but, equally clearly, my third option is *more* wrong than my second. In this respect, wrongness contrasts with obligatoriness, which doesn't appear to come in degrees. While we can surely be in a situation where we have more than one wrong option, it's much harder to see how we could be in a situation where we have more than one obligatory option.¹¹ Thus, while it makes sense to say that one action is more wrong than another, it's much less clear whether it makes sense to say that one action is more obligatory than another.¹²

¹¹ Of course, some will argue that moral dilemmas are possible and that we therefore can have more than one obligatory option. However, my general point still holds, which is simply that it's much more controversial to maintain that we can have multiple obligatory options than it is to maintain that we can have multiple wrong options.

¹² It's an interesting question whether rightness (i.e., permissibility) comes in degrees. Just as we can have more than one option that is wrong, we can have more than one option that is right. Because of this, one might think there is a symmetry between wrongness and rightness: rightness comes in degrees just as wrongness does. However, I take it that the claim that one option can be more right than another is much more contentious than the claim that one option can be more wrong than another. For an interesting discussion of the possibility of degrees of rightness, see

If wrongness comes in degrees, as it seems to, then this opens the door for a new theory of subjective rightness—namely, one that is very much like R4e, but it appeals directly to the degreed notion of objective wrongness instead of to the degreed notion of net *prima facie* rightness. However, before stating this theory precisely, I will need to set the stage a bit.

First, I will assume that each of a person's alternatives in a situation can be assigned a precise degree of objective wrongness greater than or equal to 0. For an alternative to be objectively wrong to degree 0 is for it to be not wrong at all (i.e., objectively right). The greater the degree of objective wrongness above 0 that an alternative is assigned, the more wrong it is. So, in the situation just mentioned, saving one person is objectively wrong to degree 0 (since it is objectively right); killing one person is, perhaps, objectively wrong to degree 100; and killing two people is, perhaps, objectively wrong to degree 200. Of course, it will be very difficult to actually assign precise degrees of objective wrongness to a person's alternatives in any realistic scenario. However, I shall assume that this can, in principle, be done.

Second, not only will I assume that a person's alternatives can be assigned specific degrees of objective wrongness, I will also assume that their *expected* objective wrongness can be calculated. The expected objective wrongness of an alternative is the sum, for each possible degree of objective wrongness, of this degree of wrongness times the probability that the alternative is wrong to that degree. More precisely, where A is an alternative and W_1, W_2, \dots, W_n are the possible degrees of objective wrongness of A, the

Tom Hurka's 5/4/2013 post on the topic on the PEA Soup blog (www.peasoup.typepad.com). Hurka also discusses the possibility of degrees of wrongness on the blog on 4/29/2013.

expected objective wrongness of A = {[prob(A is objectively wrong to degree W_1) \times W_1] + [prob(A is objectively wrong to degree W_2) \times W_2] + ... + [prob(A is objectively wrong to degree W_n) \times W_n]}. (Again, the sort of probability I have in mind here is epistemic probability.) It will obviously be extremely difficult to actually calculate the expected objective wrongness of a person's alternatives in any realistic situation; however, I will again assume that the expected objective wrongness of a person's alternatives can, at least in principle, be calculated.¹³

In order to better understand expected objective wrongness, it may be helpful to see how the expected objective wrongness of an alternative is calculated in a specific case. Let's return to the case of Dr. Jill. Recall that Jill knows that giving drug A will not maximize the net balance of prima facie rightness over wrongness. Thus, since Jill is an objective Rossian, she knows that this option is objectively wrong. However, she also knows that giving A will partially cure her patient, so she presumably knows that while giving A is objectively wrong, it's only mildly wrong (say, to degree 10). Let's assume that Jill's epistemic probability that giving A is objectively wrong to this degree is 1.¹⁴ Given that, the expected objective wrongness of giving A is $1 \times 10 = 10$.

With regard to giving drug B, Jill presumably knows that either it maximizes net prima facie rightness, or it has, by a large measure, the lowest net prima facie rightness

¹³ In sec. 7.7, I will discuss what my theory of subjective rightness might look like if we reject this assumption.

¹⁴ This is a simplifying assumption that helps to keep the math easy. In a more realistic version of the case, some of Jill's evidence will suggest that her action is wrong to degree 10, some of her evidence will suggest that it is wrong to degree 9.99, some will suggest that it is wrong to degree 10.01, and so on for many degrees of wrongness in the 10 ballpark.

among her alternatives. So, since she is an objective Rossian, she presumably knows that giving B either is objectively wrong to degree 0, or it is objectively wrong to some degree greater than the degree to which giving A is wrong (say, 100). Jill, however, has no idea which outcome drug B will have. So, let's assume that Jill's epistemic probability that giving B is objectively wrong to degree 0 is .5, and her epistemic probability that it is objectively wrong to degree 100 is also .5.¹⁵ Given this, the expected objective wrongness of giving B is $(.5 \times 0) + (.5 \times 100) = 50$.

The same things that I said about giving drug B apply, *mutatis mutandis*, to giving drug C. I'll assume, then, that the expected objective wrongness of this option is also 50.

The following chart summarizes all of the pertinent information needed to calculate the expected objective wrongness of Dr. Jill's options:

Table 13: Dr. Jill (expected objective wrongness)

<u>Alternative</u>	<u>Relevant possible degree of objective wrongness</u>	<u>Epistemic probability the alternative is wrong to this degree</u>	<u>Expected objective wrongness of alternative</u>
Give drug A	10	1	$10 \times 1 = 10$
Give drug B	100 0	.5 .5	$(100 \times .5) + (0 \times .5) = 50$
Give drug C	100 0	.5 .5	$(100 \times .5) + (0 \times .5) = 50$

As I've said, it intuitively seems that it is subjectively right for Jill to give drug A and subjectively wrong for her to give either of B or C. But notice that the expected

¹⁵ Another simplifying assumption for easy-math purposes.

objective wrongness of giving A is *less* than the expected objective wrongness of giving either of B or C. This suggests the following theory of subjective rightness:

SR_{MEOW}: An action A is morally right (in the subjective sense) for S(A) to perform if and only if (and because) A minimizes expected objective wrongness,

where

An action A minimizes expected objective wrongness =_{df.} S(A) can perform A, and A's expected objective wrongness is at least as small as that of any other action S(A) can perform instead.

As far as I know, no one has offered a theory of subjective rightness quite like SR_{MEOW}. However, it is an attractive one. For one thing, as we've just seen, it makes the correct conclusions about the Dr. Jill case, at least if we fill in the details of the case as I have. Note also that SR_{MEOW} will make the correct conclusions about the case on many other plausible ways of filling in its details. Suppose, for instance, that Jill is certain that giving drug A is objectively wrong to degree 15 instead of to degree 10. Then the expected objective wrongness of this option will be $15 \times 1 = 15$, which still minimizes expected objective wrongness. Or suppose that Jill doesn't know for sure how objectively wrong giving drug A is. Suppose her evidence is divided (but not equally) as to whether it is objectively wrong to degree 10 or to degree 20—so, let's suppose her epistemic probability that giving drug A is objectively wrong to degree 10 is .8, and her epistemic probability that giving drug A is objectively wrong to degree 20 is .2. Then the expected objective wrongness of this option is $(10 \times .8) + (20 \times .2) = 12$, which again continues to minimize expected objective wrongness. The point is, SR_{MEOW} will make

the intuitively correct conclusions about the Dr. Jill case on a great many ways of precisifying it. It therefore avoids the problems of R4.

SR_{MEOW} also avoids the problems of R4e. R4e, recall, ties subjective rightness to expected net prima facie rightness. The problem with this, we've seen, is that it is customized too closely to those who accept Rossianism about objective rightness; intuitively, a theory of subjective rightness should be relativized to the theory of objective rightness that an agent is justified in accepting, whether that theory is a Rossian theory or a non-Rossian one. SR_{MEOW}, however, is fully relativized in this way. According to it, the subjective rightness of an action is determined by the agent's evidence concerning its possible degrees of objective wrongness. The theory makes no substantive assumptions about what in fact determines objective wrongness. Rather, the theory allows agents to determine this on their own, and it makes the subjective rightness of their actions a function of whatever theory of objective rightness/wrongness they are justified in accepting.

Finally, SR_{MEOW} does not need to resort to radical disjunctivization in order to relativize the subjective rightness of an action to the theory of objective rightness the agent is justified in accepting. For this reason, SR_{MEOW} avoids the problem of R4ed, which is another mark in its favor.¹⁶

¹⁶ Note that SR_{MEOW} also seems to make the correct conclusion about the feverish child case discussed at the beginning of this chapter. Recall that in that case, a mother can give her sick child Advil or Tylenol. She thinks—and, we can presume, is fully justified in thinking—that either drug would be fine to give to her child. So, plausibly, she believes—and is justified in believing—that each option is objectively right. The expected objective wrongness of each of her options is therefore $0 \times 1 = 0$, and each option is subjectively right according to SR_{MEOW}, just as it intuitively seems. The same thing will hold on other ways of making the case more concrete.

In sum, SR_{MEOW} solves all of the problems of its predecessors. For this reason, it is an appealing theory, one that deserves to be taken seriously. However, in order to make the appeal of SR_{MEOW} even more apparent, I'd now like to compare it to two additional theories of subjective rightness that have been offered recently: Holly Smith's (2010) and Fred Feldman's (2012). Smith and Feldman agree with me about many things. They both think that we should draw a distinction between objective and subjective rightness. They both think that the Dr. Jill case is an important test case for a theory of subjective rightness. They both agree that the move to expected utility (or, presumably, to expected net prima facie rightness) cannot adequately handle this case. And they both maintain that subjective rightness should be relativized to the objective moral theory of an agent. However, while Smith and Feldman share my general approach, the theories of subjective rightness they end up endorsing are very different from mine. Smith's and Feldman's theories are therefore two of SR_{MEOW} 's most important competitors. For this reason, I will now discuss them.

7.4. Smith and Feldman

I'll begin with Smith's (2010) theory. Smith thinks that in order to give a theory of subjective rightness, we first need to talk about moral principles. In particular, we need to say what it is for a moral principle to be *objective* and what it is for a moral principle to be *subjective* (92). Smith starts with objectivity. She gives a precise account of what it is for a moral principle to be objective (93), but the details don't matter much for my purposes. It will suffice for me to note that principles like "an action is objectively right

iff it maximizes utility” and “an action is objectively right iff its maxim is universalizable” count as objective moral principles on Smith’s account.

After explaining what it is for a moral principle to be objective, Smith then explains what it is for a moral principle to be subjective. According to Smith, a subjective moral principle is always relativized to an objective one. She seems to think that for any objective moral principle, there will be a set of subjective moral principles that correspond to it. Each of the principles in this set is, Smith suggests, “appropriate” to the objective one (2010, 94, 95-97, 99, 102, 106). This set will be quite large, possibly infinite. Smith gives a complicated account of what it is for a moral principle to be subjective (93-94), but for my purposes, there is again no need to discuss the details. It will suffice for me to say that, relative to the utilitarian principle of objective rightness mentioned in the previous paragraph, the following principles of subjective rightness will plausibly be appropriate to it: an action is subjectively right if it maximizes utility, an action is subjectively right if it maximizes expected utility, and an action is subjectively right if it maximizes minimum utility.

Smith also notes that for most sets of subjective moral principles, some of the principles in the set will be action-guiding, while others won’t. Which of these principles will be action-guiding for the agent greatly depends on what the agent is like and what situation she is in. The action-guidingness of a subjective moral principle is therefore relativized to agents and their circumstances. Smith (2010, 101-2) says a great deal about precisely what it is for a moral principle to be action-guiding, but I will set aside the details for now (I will return to them in the next section). For now, I will simply assume that the notion of a moral principle’s being action-guiding is well-enough understood.

Finally, Smith suggests that for any set of subjective principles, there will be a “hierarchy” of the principles within it (2010, 92, 94, 99, 102). The idea here seems to be this. As we’ve seen, Smith holds that for each objective moral principle, there is a corresponding set of subjective moral principles that is appropriate to it. And Smith’s idea seems to be that some of the principles in this set are “higher up,” “more important,” or “more appropriate” to the objective principle than others are. Consider, for example, a utilitarian theory of objective rightness. As I’ve indicated, the set of subjective moral principles appropriate to it plausibly includes principles like (1) an action is subjectively right if it maximizes utility, (2) an action is subjectively right if it maximizes expected utility, and (3) an action is subjectively right if it maximizes minimum utility. (1) is plausibly at the top of the hierarchy of principles in this set—it is the “most important” one, or the “most appropriate” to objective utilitarianism. And (2) is plausibly “higher up” (“more important,” “more appropriate”) than (3) is.

So far, I’ve merely been discussing the nature of objective and subjective moral principles, along with various features of them. But, I’m now in a position to say what, according to Smith, it is for an action to be subjectively right. Her theory can be stated as follows:

SR_{HS}: An action A is subjectively right for a person to perform (relative to her theory of objective rightness T) if and only if A is prescribed by the highest action-guiding subjective moral principle in the hierarchy of moral principles that is appropriate for T (see 2010, 100).

Here, then, is a decision procedure we can use to determine which of our alternatives are, according to Smith, subjectively right (relative to our theory of objective rightness):

1. Determine what our theory of objective rightness is.

2. Figure out all of the subjective moral principles that are appropriate to it.
3. Figure out which of these subjective principles are action-guiding.
4. Determine the hierarchy of importance among the action-guiding ones.
5. Locate the highest principle in the hierarchy.
6. Use it to determine our subjectively right alternatives (we should easily be able to do this since the principle will necessarily be action-guiding).

One might wonder whether we could ever actually complete these steps, and thus whether Smith's theory of subjective rightness really is action-guiding, as she claims any theory of subjective rightness must be (steps 2 and 4 seem particularly difficult to complete). However, I won't press this point here. The objection I want to press has to do with her use of the concept of appropriateness and her reliance on a hierarchy of subjective principles.

As we've seen, Smith suggests that for each principle of objective rightness, there is a set of subjective principles that is "appropriate" to it. However, the concept of appropriateness at issue here is a bit mysterious. Clearly the principle that an action is subjectively right if it maximizes utility is appropriate to a principle of objective utilitarianism. So is the principle that an action is subjectively right if it maximizes expected utility. But the set of subjective moral principles appropriate to objective utilitarianism (or any other objective moral principle) needs to be *enormous*—otherwise Smith won't be able to guarantee that for any agent in any circumstance, there will be a subjective moral principle appropriate to her objective principle that is action-guiding. But it's hard for me to imagine a sense of "appropriate" that will generate these massive sets of subjective principles.

To further press the issue, note that Smith suggests that for any set of subjective principles, there will be a "bottom level" principle like this one: an action is subjectively

right if the agent believes she can perform it (2010, 109n73). This is the subjective principle that agents may need to use in circumstances of extreme ignorance. But is this principle really “appropriate” to any objective principle? Is it appropriate to objective utilitarianism, for instance? It’s hard to see what sense of “appropriate” will allow us to say “yes.” So, I think one should worry about Smith’s use of appropriateness in her theory of subjective rightness. The notion does not seem much clearer than the notion of subjective rightness itself. Because of this, giving an account of subjective rightness in terms of appropriateness has little pay-off; Smith merely seems to be substituting one mysterious notion for another.

Even more troubling, I think, is Smith’s reliance on a hierarchy of subjective moral principles. Recall: Smith thinks that for any set of subjective principles, some will be “higher up,” “more important,” or “more appropriate” to the associated objective principle than others will be. But, of course, one will wonder: what makes one subjective principle “higher up” (“more appropriate,” etc.) than another? This is a vexing question, and one that Smith does not attempt to answer. She sets it aside for another occasion (2010, 92-93). This isn’t necessarily bad—one can’t do everything in a single philosophy paper. However, it *is* an important question, and one, I think, that is going to be very difficult to answer. I suspect that giving an account of what makes one subjective principle “higher up” than another in a hierarchy is going to be just as difficult as giving an account of subjective rightness in the first place. So, again, one has to wonder how much has been gained by Smith’s theory.¹⁷

¹⁷ Feldman (2012, 156) suggests a way for Smith to make sense of the idea of one subjective principle’s being higher up in the hierarchy than another. Smith could say that one subjective

Feldman's (2012) recent theory of subjective rightness suffers from a similar difficulty. Feldman focuses, at least initially, on giving an account of subjective rightness for those who accept utilitarianism about objective rightness. According to his view, an action is subjectively right for such a person iff it is an outcome of the Utilitarian Decision Procedure. In order to determine whether an alternative of ours is the outcome of the Utilitarian Decision Procedure, Feldman says that we need to follow these steps:

1. First, we need to consider all of the actions we think we can perform (described in action-guiding ways).
2. Next, we need to consider what we take the values of these actions to be.
3. Then, we need to take into consideration our views about the morality of risk as it applies to our situation.
4. Finally, we need to identify the actions that "seem most nearly consistent with the general policy of maximizing utility where possible while avoiding things that put people at excessive risk of serious harm."
5. The actions identified in (4) are the "outcome" of the Utilitarian Decision Procedure. These actions are the ones that are subjectively right for us to perform, at least if we are utilitarians about objective rightness (166-67).

While the Utilitarian Decision Procedure only tells us which actions are subjectively right for those who accept utilitarianism about objective rightness, Feldman notes that similar procedures can be constructed to tell us which actions are subjectively right for agents who accept other views of objective rightness (2012, 167-71). Consider, for instance, a virtue ethicist who holds that an action is objectively right iff it maximizes the virtue/vice quotient (an action maximizes the virtue/vice quotient iff the balance of

principle, S1, is higher up in the hierarchy than another, S2, iff S1 has a higher "success rate" than S2, where the success rate of a subjective principle is the percentage of cases in which the action recommended by the principle is the same as the action recommended by the objective moral principle that the subjective principles are associated with. Feldman argues, however, that on this proposal, Smith's theory of subjective rightness fails to be action-guiding.

virtue over vice that the action manifests is at least as great as that of any of its alternatives). According to Feldman, the actions that are subjectively right for proponents of this view are the ones that are the outcome of the Virtue Ethics Decision Procedure, which looks much like the Utilitarian Decision Procedure except that talk of “value” is replaced with talk of “virtue” and “vice,” as appropriate. The same thing goes, *mutatis mutandis*, for those who accept Kantianism about objective rightness, or Rossianism, or a Rights Theory, etc. Given this, it seems we can state Feldman’s theory in a more general way as follows:

SR_{FF}: An action is subjectively right if and only if it is an outcome of the decision procedure associated with the theory of objective rightness that the agent accepts,

where

The decision procedure associated with the theory of objective rightness that an agent accepts looks relevantly like the Utilitarian Decision Procedure above.

One issue that arises at this point is this. What if there is no theory of objective rightness that an agent accepts? Perhaps the agent simply hasn’t thought much about right and wrong. She’s never taken an ethics class. She doesn’t know what utilitarianism is, or Kantianism, or virtue ethics. She may have some opinions or intuitions about the moral status of some actions in real life cases, but it would be difficult to maintain that there is any theory of objective rightness that she accepts. What would Feldman say then?¹⁸

¹⁸ A similar issue arises for Smith’s theory. In conversation, Feldman has indicated that he would say that in the case at hand, the agent’s alternatives have no subjective moral status (i.e., they are neither subjectively right nor subjectively wrong).

Another potential problem for Feldman's view—and one that I think is more serious—has to do with the notion of consistency that appears in the Utilitarian Decision Procedure (and presumably in the decision procedures associated with every other theory of objective rightness). Feldman claims that, for those who accept utilitarianism about objective rightness, subjectively right actions are ones that seem most “consistent” with the policy of maximizing utility while avoiding putting people at serious risk of harm. But what exactly does Feldman mean by “consistent” in this context?

Consider the Dr. Jill case, for instance. In the version of the case that Feldman discusses, Jill is a utilitarian about objective rightness, not a Rossian about it, and she is certain that drug A will have pretty good consequences (since it will partially cure her patient, John), and she is certain that one of drug B or C will have the best consequences and the other will have the worst (since one will cure John completely and the other will kill him), but she doesn't know which is which. As Feldman imagines the case, Jill “thinks it would be morally wrong in this particular case to put John at serious risk of death unless it is absolutely necessary to save his life” (2012, 161). Because of this, Feldman thinks that a Utilitarian Moral Guide will recommend that Jill give her patient the safe drug—in other words, giving him this drug will be outcome of the Utilitarian Decision Procedure. Feldman therefore seems to hold that giving this drug is most consistent with Jill's view about objective rightness and her views about the morality of the risks involved in the situation.

But the sort of consistency that Feldman has in mind here surely isn't the logical sort. While giving the safe drug seems logically consistent with Jill's objective moral viewpoint and her views about risk, so does giving either of the other drugs; Jill's

objective moral viewpoint and her views about risk do not seem to *entail* that she give the safe drug rather than a risky one. But if the notion of consistency that Feldman has in mind isn't logical consistency, then what is it? As far as I can tell, Feldman never addresses this question, though I suspect that he ultimately has in mind some notion of appropriateness, or fittingness, or something along these lines. Feldman's view of subjective rightness therefore seems to have a problem much like the one that Smith's has: it makes use of a somewhat opaque normative notion that may appear at least as mysterious as the notion of subjective rightness itself. It's therefore unclear how much ground has been gained.

Notice, however, that SR_{MEOW} avoids this problem. The theory makes use of normative concepts that, I am assuming, are readily understood. The central normative concept it appeals to is the concept of objective wrongness, and this concept is already a fixture in the current ethical landscape. It therefore seems to me to be preferable to give a theory of subjective rightness in terms of objective wrongness instead of in terms of a normative concept like Smith's appropriateness or Feldman's consistency, or some other normative concept that is difficult to get a grip on. And even if one comes to doubt the intelligibility of the concept of objective rightness/wrongness, note that Smith's and Feldman's theory make use of it too (in addition to the other, vaguer normative concepts I have highlighted). So, attacking the coherence of objective rightness will undermine their theories as well as mine. The advantage of the theory I've offered is that it doesn't employ a (mysterious) normative concept in addition to objective rightness/wrongness. My point is that once the intelligibility of objective rightness/wrongness is granted, we

have all the resources we need to construct a plausible theory of subjective rightness. For these reasons, I think the theory I have offered is preferable to Smith's and Feldman's.

At this point, however, I suspect that Smith and Feldman would argue that SR_{MEOW} has a glaring problem: it's not action-guiding. This feature of the theory, they might insist, overshadows all of its attractions, and it should therefore be rejected. In the next section, I will address this objection.

7.5. SR_{MEOW} and Action-Guidance

In order to determine whether SR_{MEOW} indeed fails to be action-guiding (and whether this is a problem), we first need to get clear about what it is for a moral theory to be action-guiding. In fact, this is a rather difficult thing to get clear about since philosophers seem to mean different things when they say that a moral theory is (or is not) action-guiding. However, I will simply follow Holly Smith on these matters since she has, perhaps more than anyone else, attempted to clarify the concept of action-guidance (see her 1988; 2010; 2012). As we've seen, Smith thinks that the subjective principles that determine the subjective rightness of one's actions must be action-guiding (or, as she says, "guidance adequate"). Smith initially suggests that for a principle or theory of subjective rightness to be action-guiding is for it to be one "from which agents can derive moral guidance in every situation in which they find themselves, even though an agent may be uncertain or mistaken about which actions have the features that would make them objectively right in that situation" (2010, 72).

However, as Smith (2010, 72n14, 101) notes, this can't be exactly right. Consider the following (admittedly implausible) theory of subjective rightness: an action is

subjectively right iff it maximizes utility. Notice that we *can* derive moral guidance from this theory in every situation in which we find ourselves. In any scenario we encounter, we can use the theory to tell us that all of our options that maximize utility are subjectively right, and all of the ones that don't are subjectively wrong. So, this theory appears to count as an action-guiding one on Smith's initial proposal. But surely we do not want to say this; the utilitarian theory in question will be regarded by most philosophers as a paradigmatic example of a theory that *fails* to be action-guiding.

As Eugene Bales points out, utilitarian theories, such as the one mentioned in the previous paragraph, seem to be useless as a guide to action because although an agent can use them to derive the moral status of an option when it is described as being a "maximizer of utility" or a "non-maximizer of utility," these descriptions are unhelpful since the agent will typically not know whether the option can be correctly described in either of these ways (1971, 264; see also H. Smith 2010, 101). Thus, one plausible way of modifying Smith's initial proposal in light of the present problem is to say that for a theory of subjective rightness to be action-guiding is for agents to be able to use it in any situation they face to derive the subjective moral status of their options, *when their options are stated in helpful ways* (see H. Smith 2010, 100-2).¹⁹ Of course, if we adopt this proposal, we'll need to specify what it is for an act-description to be "helpful." Fortunately, Smith rises to the occasion. However, her account of helpful act-descriptions is highly complex, and we can safely ignore the details here. For our

¹⁹ Feldman (2012, 153-56) also appears to understand the action-guidingness of a subjective moral theory in roughly this way, though he prefers to use the term "implementability" instead of "action-guidingness."

purposes, the notion of the action-guidingness of a subjective moral theory should be clear enough (I hope!) to forge ahead.

Turning now to the charge that SR_{MEOW} fails to be action-guiding, I think it should be fairly clear that SR_{MEOW} is guilty as charged. In order for an agent to be able to use SR_{MEOW} to derive the subjective moral status of her alternatives, she will presumably need to know (a) what her alternatives are, (b) what degrees of objective wrongness they could have, and (c) precisely how justified she is in believing that each of her alternatives is wrong to each of these possible degrees. Agents, however, will typically not know these things (especially (c)). And even if an agent could come to know them, she will still have a multitude of mathematical calculations to perform before she can know which of her alternatives minimize expected objective wrongness—in particular, for each of her alternatives, she'll have to multiply each of its possible degrees of wrongness by the amount to which she is justified in believing that the alternative is wrong to that degree and then add up the products. In any ordinary scenario, agents will not be able to make all of these calculations (imagine the time it would take!). So, I am willing to concede that agents will rarely be able to use SR_{MEOW} to derive the subjective moral status of their alternatives; the theory therefore fails to be action-guiding.

However, I suspect that every theory of subjective rightness—or, more carefully, every *plausible* one—will fail to be action-guiding. I've already suggested why Smith's own theory is not action-guiding. As mentioned above, in order to use her theory to derive the subjective moral status of an action, an agent will need to know (a) what theory of objective rightness she accepts, (b) what principles of subjective rightness are appropriate to it, (c) which of these principles are action-guiding, (d) and which of these

action-guiding principles is the highest in the hierarchy of importance. But agents will sometimes—perhaps often—not know these things (especially (b) and (d)), and so Smith’s theory will not be action-guiding.

Similar things can be said of Feldman’s theory. In order to use his theory to determine the subjective moral status of an action, an agent will need to know (a) what theory of objective rightness she accepts, (b) what alternatives she thinks she can perform, (c) what her view about the morality of risk is as it pertains to her situation, and (d) which perceived alternatives seem to her most nearly consistent with her theory of objective rightness and her views about risk. But I think that agents will sometimes be in situations where they do not know at least some of these things (especially (c) and (d)), and so Feldman’s theory will not be action-guiding either.

However, Smith and Feldman are, perhaps more than any other theorist, committed to providing an account of subjective rightness that is action-guiding. So, if even *their* theories fail to be action-guiding, then I suspect that no plausible theory of subjective rightness will be. And if I’m right about this, then that would give us good reason not to demand action-guidance from a theory of subjective rightness.

But I don’t want my only response to the charge in question to be a *tu quoque* one. A better response, I think, would be to admit that SR_{MEOW} is not action-guiding but deny that this is a problem. To see why SR_{MEOW}’s lack of action-guidance is not a problem, it’s important to understand the nature of the project I take myself to be engaged in. As I’ve suggested, there’s an intuitive sense of “right” according to which it is morally right for the mother in the feverish child case to give her child the Tylenol (which happens to kill the child), and it is morally right for Dr. Jill to give her patient the safe drug (which

she knows will not maximize net prima facie rightness). I think this sense of “right” is an interesting and important one, and it’s the one that I have been concerned with in this chapter—I have referred to it as the “subjective” sense of “right.” Of course, like any important concept, it is one about which philosophers are bound to wonder. One thing philosophers will like to know is what makes subjectively right actions subjectively right. Why is it subjectively right for Dr. Jill to give the safe drug? Why is it subjectively right for the mother to give the deadly Tylenol? What explains why all subjectively right actions like these ones are subjectively right? These are interesting philosophical questions, and they are the ones that I’ve been chiefly concerned with here. The theory I have proposed, SR_{MEOW} , answers these questions, and it answers them correctly (at least in my mind). Why, then, would it matter if SR_{MEOW} fails to be action-guiding? Whether or not SR_{MEOW} is action-guiding seems to me to be irrelevant. After all, its aim is not to guide actions; its aim is, rather, to specify the feature in virtue of which all subjectively right actions are subjectively right. And that’s precisely what it does. So what if it doesn’t do more?

What I’m saying here resembles what R. Eugene Bales says about act utilitarianism in his well-known paper “Act-Utilitarianism: Account of Right-Making Characteristics or Decision-Making Procedure?” (1971). In that paper, Bales responds to the charge that act utilitarianism is problematic since it fails to provide us with a useful decision-making procedure. Bales argues that act utilitarianism is merely trying to identify the necessary and sufficient conditions for the moral rightness of actions. Its purpose is not to provide us with a decision-making procedure that we can easily use to determine what we should do. Because of this, Bales argues, it is illegitimate to object to the theory by pointing out

that it fails to provide us with a useful decision-making procedure, for that is not what it's trying to accomplish.²⁰ I am saying something similar about SR_{MEOW}. As I see it, SR_{MEOW}, like act utilitarianism, is merely trying to identify (an explanatory set of) necessary and sufficient conditions for the moral rightness of actions. However, while act utilitarianism is trying to identify the conditions of *objectively* morally right actions, SR_{MEOW} is trying to identify the conditions of *subjectively* morally right actions. The central concern of SR_{MEOW} is to identify these conditions, not to provide us with an easily implementable decision-making procedure. One therefore cannot object to the theory merely by pointing out that it fails to be action-guiding, for action-guiding is not something it aspires to be.

At this point, I can imagine an objector saying something like the following. While it's okay to hold that a theory of objective rightness merely provides us with an account of right-making characteristics and not a useful decision-making procedure, it's unacceptable to say a similar thing about a theory of subjective rightness. After all, isn't the whole point of a theory of subjective rightness to give us moral guidance in situations of moral uncertainty? Our theory of objective rightness—no matter whether it's a

²⁰ Bales's argument is anticipated by Ross. In chapter 5 of *FE*, Ross responds to some objections raised by W. A. Pickard-Cambridge. One of these objections is that Ross's view is problematic because it is difficult to apply—Ideal Utilitarianism is much easier to apply and is thus, in Pickard-Cambridge's opinion, superior to Ross's theory. In response to this objection, Ross says, "My main answer, however, would be...that any appeal to the ease of applying one view or the other is beside the mark. It is not the business of moral philosophy to provide us with a theory which is easy to apply. Its business, or the part of its business with which we are at present concerned, is to say on what the rightness or wrongness of actions in fact depends. The fact that it would be easier to recognize our duty if it depended on factor *a* only than it would be if it depended on factors *a*, *b*, and *c* has no tendency whatever to prove that in fact it depends on *a* alone" (1939, 90-91).

utilitarian theory, a Kantian theory, a contractualist theory, or some other theory—will almost surely fail to be action-guiding. It thus appears that our theory of objective rightness will need supplementation. And isn't our theory of subjective rightness supposed to provide us with that supplementation?

In response to this, I readily admit that our theory of objective rightness will likely be useless as a guide to action and that we will therefore need to look elsewhere to obtain guidance in situations of moral uncertainty. I also admit that searching for a useful decision-making procedure to supplement our theory of objective rightness with is a worthy project. However, I would stress that this is not the project that I have been concerned with. As I've said, I'm merely trying to determine the conditions under which an action is subjectively right. I think this project is a worthy one as well. I admit that it would be nice if, in specifying these conditions, we could solve other problems—it would be nice if, for instance, we could specify these conditions in an action-guiding way (just as it would be nice if we could specify the conditions of *objective* rightness in an action-guiding way). But I see no reason to insist that these conditions cannot be correctly specified unless they simultaneously solve these other problems. To be sure, someone may be able to find a way of specifying these conditions in an action-guiding way, and perhaps such a theory would, for that reason, be preferable to the one I've offered here. However, I know of no such theory. So, until one can be found—and I doubt that there is one to find—we must be content to divide and conquer. We must be content, that is, to search for the ground of subjective rightness and a useful decision-making procedure independently. And if what I've argued here is correct, then we've made substantial progress in our search for the former. And that, I think, is a significant result.

7.6. A Problem Involving Supererogation

The next objection to SR_{MEOW} that I'd like to consider has to do with supererogatory actions. An agent might find herself in a situation where she is justified in believing that one or more of her options is not only not objectively wrong, but is also objectively supererogatory. In some of these situations, SR_{MEOW} may appear to deliver unwelcome results.

To see this, let's look first at another case involving a fire, one that is inspired by Kagan (1989, 16, 240) and Portmore (2011, 138-41) (recall that I also discussed a fire case in sec. 6.7):

Fire (Second Version)

A house is burning down and a child is trapped inside. I'm the only one who can help—the firefighters won't arrive in time. I have three options. First option: do nothing. I know that if I take this option, I will be safe, but the child trapped in the house will die. Second option: run into the house and save the child. I know that if I perform this option, the child will be saved, but I will sustain severe burns that will leave me permanently disabled. Third option: run into the house and save the bed the child is hiding under. I know that if I do this, the bed will be unscathed, but the child will die and I will still sustain severe burns that will leave me permanently disabled.

What is the objective moral status of my options? I think that many will claim that it is objectively permissible for me to do nothing. I think many will also say that it's objectively permissible (indeed it's supererogatory) for me to run into the house and save the child. But what about running into the house and saving the bed? The matter here is

more complex; however, I suspect that many will say that doing that is objectively *wrong*: once I have made the decision to enter the house and sacrifice my own well-being, I am obliged to save the child rather than the bed—after all, saving the child will result in no greater cost to me.²¹

One might challenge any or all of these claims about the objective moral status of my options, but let's assume that I, at least, am justified in believing them. That is, suppose that I am fully justified in believing that it's not objectively wrong (i.e., it's objectively right) for me to either do nothing or enter the house and save the child, but it is objectively wrong (say, to degree 20) for me to enter the house and save the bed. Given this information, the expected objective wrongness of my options is calculated as follows:

Table 14: Fire (Second Version)

<u>Alternative</u>	<u>Relevant possible degree of objective wrongness</u>	<u>Epistemic probability the alternative is wrong to this degree</u>	<u>Expected objective wrongness of alternative</u>
Do nothing	0	1	$0 \times 1 = 0$
Save the child	0	1	$0 \times 1 = 0$
Save the bed	20	1	$20 \times 1 = 20$

As can be seen from this chart, doing nothing and saving the child each minimizes expected objective wrongness, but saving the bed does not. SR_{MEOW} therefore implies that the former two options are subjectively right while the latter option is subjectively wrong. These seem to me to be the correct conclusions to make about the case.

²¹ See, e.g., Kagan (1989, 16, 240) and Portmore (2011, 139-40).

However, it may seem that SR_{MEOW} runs into a problem when we consider the following variant of the case:²²

Fire (Third Version)

Everything is the same as Fire (Second Version) except that I now have only two options. First option: do nothing. Again, I know that if I do this, the child in the house will perish but I will remain safe. Suppose I am fully justified in believing that this option is objectively permissible. Second option: enter the house on a dangerous rescue mission. I know that if I take this option, I might either choose to save the child, or I might choose to save the bed instead; either way, I know I will sustain severe burns that will leave me permanently disabled. Suppose I also know that if I do end up saving the child, my mission will be objectively right, but if I end up saving the bed, then my mission will be objectively wrong (to degree 20). But I'm just not sure which thing I will end up saving: my evidence is equally divided on the matter. So, then, let's suppose that I am justified in believing to degree .5 that my second option is objectively right, and I am justified in believing to degree .5 that it is objectively wrong (to degree 20).

Given the information stated in this case, the expected objective wrongness of my options is calculated as follows:

²² Thanks to Pete Graham for pushing me to consider this type of case.

Table 15: Fire (Third Version)

<u>Alternative</u>	<u>Relevant possible degree of objective wrongness</u>	<u>Epistemic probability the alternative is wrong to this degree</u>	<u>Expected objective wrongness of alternative</u>
Do nothing	0	1	$0 \times 1 = 0$
Enter house on rescue mission	0 20	.5 .5	$(0 \times .5) + (20 \times .5) = 10$

As the chart demonstrates, doing nothing minimizes expected objective wrongness, but entering the house does not. SR_{MEOW} therefore implies that doing nothing is my only subjectively permissible option: entering the house on the rescue mission is subjectively wrong.

Is this a problem? I'm not sure. After all, I am in a situation where I have an option that I know is objectively right. My other option might be objectively right, but, then again, it might be objectively wrong. Because of this, it might seem that I am in fact subjectively obligated to perform the first option: why should I risk acting wrongly when I can do something that is surely right? So, I can imagine someone simply accepting SR_{MEOW} 's implications about this case. However, I can also imagine someone finding SR_{MEOW} 's implications about this case to be counterintuitive. I can imagine someone insisting that each of my options is subjectively permissible. However, as I will now explain, there is a relatively easy way of modifying SR_{MEOW} so that it generates this conclusion.

Earlier, I said that each of a person's alternatives in a situation can be assigned a specific degree of objective wrongness greater than or equal to 0. However, let me now amend this proposal so that *negative* degrees of objective wrongness are possible. If an

action is objectively wrong to a negative degree, this indicates that not only is it objectively not wrong, but it is also objectively *supererogatory*. The greater the negative degree of objective wrongness an action is assigned, the more objectively supererogatory it is.²³ Actions that are objectively wrong to degree 0 are not objectively wrong, but neither are they supererogatory. And, as before, actions that are objectively wrong to a positive degree are objectively wrong—the higher the positive degree of objective wrongness an action is assigned, the more wrong it is.

Let's now return to Fire (Third Version). In the case, I am presumably justified in believing that if I enter the house and end up saving a child, then my action will not only be objectively right, but it will also be objectively supererogatory. So, let's suppose that I am justified in believing that in those circumstances, my action will be wrong to some *negative* degree (say, -50). Given this, we can update the above chart as follows:

Table 16: Fire (Third Version), updated

<u>Alternative</u>	<u>Relevant possible degree of objective wrongness</u>	<u>Epistemic probability the alternative is wrong to this degree</u>	<u>Expected objective wrongness of alternative</u>
Do nothing	0	1	$0 \times 1 = 0$
Enter house on rescue mission	-50 20	.5 .5	$(-50 \times .5) + (20 \times .5) = -15$

Since entering the house now minimizes expected objective wrongness, SR_{MEOW} implies that it is subjectively right. However, the theory now implies that doing nothing is subjectively wrong since it no longer minimizes expected objected wrongness. We

²³ Although it is controversial whether rightness comes in degrees (see note 12), supererogatoriness surely does.

therefore haven't yet achieved the result we are after, which is to make it so that *each* of my options is subjectively right.

What we need to do, I think, is modify SR_{MEOW} slightly. Consider the following view:

SR_{MEOW+} : An action A is morally right (in the subjective sense) for S(A) to perform if and only if (and because) *either* (a) A minimizes expected objective wrongness, *or* (b) A's expected objective wrongness is less than or equal to 0.

Since the expected objective wrongness of each of my options in Fire (Third Version) is less than 0, SR_{MEOW+} implies that each is subjectively right. SR_{MEOW+} therefore makes the intended conclusions about the case. What's more, the theory is an attractive one. Clause (b) says that any action that is expected to be objectively right (or supererogatory) is also subjectively right. On its face, that seems plausible. Of course, we will not always be in a situation where we expect each of our options to be objectively right (or supererogatory). Clause (a) states that in those cases, our subjectively right actions are the ones that minimize expected objective wrongness. Given the main thrust of my argument in this chapter, that should seem plausible as well. SR_{MEOW+} therefore seems to me to be a promising way of modifying SR_{MEOW} in light of Fire (Third Version), though let me reiterate that it's unclear to me whether the theory ever needed modification in the first place.

7.7. The Imprecise Justificatory State Problem

Allow me to set aside the complexities raised in the previous section and return to SR_{MEOW} . There is one additional problem for it that I'd like to address, a problem that

arises if our justificatory state is not as precise as I have been assuming. As we've seen, the expected objective wrongness of an alternative is the sum, for each possible degree of objective wrongness, of this degree of wrongness times *the* probability that the alternative is objectively wrong to that degree. The probability I have in mind here is epistemic probability, which has to do with degrees of justified belief or degrees of evidential support. More precisely, then, the expected objective wrongness of an alternative is the sum, for each possible degree of objective wrongness, D, of this degree of wrongness times *the* degree to which the agent is justified in believing that the alternative is objectively wrong to degree D. Note the italicized word in the previous sentence: in order to calculate the expected objective wrongness of an alternative, there must be, for each possible degree of objective wrongness, D, some precise degree to which the agent is justified in believing that the alternative is objectively wrong to degree D. But, one might think, that is absurd. Consider, for instance, the degree of objective wrongness of 14.5938547. To what degree am I justified in believing that one of my alternatives is wrong to this degree? It's very tempting to think that there is no determinate answer to this question: there is no such degree of belief. At most, there is an *interval* or *span* of degrees of belief to which I am justified in believing this. But, the thought goes, similar things can be said about virtually every other precise degree of objective wrongness and about virtually every other person. And if that is the case, then it looks like it will often (perhaps always) be impossible to calculate the expected objective wrongness of a person's alternatives, at least if expected objective wrongness is calculated as I've indicated.

Here's one way of responding to this worry. One might admit that we can't *know* the precise degree to which I am justified in believing that an alternative of mine is objectively wrong to degree 14.5938547. Our imperfect intellectual capacity makes it impossible for me, or you, or any other human to know this. Because of this, it's very easy to conclude that there simply is no such degree. But, of course, this line of reasoning is fallacious. The fact that I can't know the precise degree to which I am justified in believing that an alternative of mine is objectively wrong to degree 14.5938547 doesn't imply that there isn't such a degree.

I am sympathetic with this line of response. However, I recognize that many will not be. Many will insist that *God Himself* wouldn't be able to determine the precise degree to which I am justified in believing that an alternative of mine is objectively wrong to degree 14.5938547. And the reason for this is not that God has some intellectual defect or shortcoming (that would be incoherent); rather, it's because there simply is no such degree to be known.

Since I suspect that some will likely hold this view, what I'd like to do now is sketch a way of modifying SR_{MEOW} to respond to it. I hope to show that a version of the theory can survive even if it turns out that our justificatory state is imprecise in the way just described.²⁴

²⁴ There is now a burgeoning literature on the topic of whether our credences are always (or ever) sharp, whether our evidence always (or ever) justifies us in believing a proposition to a precise degree, and how one's view about these matters impacts Bayesian decision theory (see, e.g., Joyce 2010; Elga 2010; White 2010). The remainder of this section has been informed by this literature. Because this literature is vast, and because it tends to be highly complex, it's unclear to me the extent to which the ideas contained in this section are original. However, it's worth noting that this literature typically concerns epistemic or prudential rationality, not subjective moral

First, I'd like to restate the problem in a slightly different way. Earlier, when I explained how to calculate the expected objective wrongness of an alternative, it was clear that I was presuming that for any agent, his or her justificatory state can be represented by a completely determinative function that takes as an input any proposition and outputs a sharp degree of belief that the agent is justified in having toward that proposition (call such a function a "justificatory state function"). But the worry is that the justificatory state of most agents—perhaps *every* actual agent—cannot be represented in this way. At most, the justificatory state of an agent can be represented by a *set* of justificatory state functions.

Here's a more picturesque way of thinking about this. Consider an extremely opinionated politician. He has beliefs about everything. Ask him any question, and he'll give you a definite (though not necessarily correct) answer. So, if you ask him how justified you are in believing that one of your alternatives is objectively wrong to degree 14.5938547, our politician will have an answer. ".6947," he might say. Now, the claim currently under consideration is that it is impossible to represent the justificatory state of most agents with any single opinionated politician. To represent the justificatory state of most agents, we'll need to appeal to a whole *congress* of such politicians (and this congress will typically be quite large). So, then, how justified are you in believing that one of your alternatives is objectively wrong to degree 14.5938547? Perhaps one of the opinionated politicians who partially represents your justificatory state will say ".6948."

rightness, and for this reason doesn't address precisely the same topic that I have been discussing. Perhaps, then, my chief claim to originality in this section is in my application of some lessons gleaned from Bayesian literature to a new area of inquiry.

Perhaps another will say “.6946.” Another might say “.6945.” And so forth. This captures the intuitive idea that there is a *range* or *set* of degrees to which you are justified in believing that one of your alternatives is objectively wrong to some precise degree.²⁵

If an agent’s justificatory state really cannot be represented by only one justificatory state function, then SR_{MEOW} will need to be modified. However, it strikes me that there is a natural way of doing this. We are currently supposing that if we are going to represent an agent’s justificatory state with a justificatory state function, we can do so only by appealing to many such functions. But recall that each justificatory state function that we use to represent an agent’s justificatory state is, by definition, completely determinative: each takes as an input any proposition and outputs a sharp degree of belief that the agent is justified in having toward that proposition. So, we *can* calculate the expected objective wrongness of an agent’s alternatives for each of the justificatory state functions that represent her justificatory state. We can, then, modify SR_{MEOW} as follows:

SR_{MEOW}' : An action A is morally right (in the subjective sense) for S(A) to perform if and only if (and because) A minimizes expected objective wrongness on each of the justificatory state functions that represent the agent’s justificatory state.

In more picturesque terms: an action is subjectively right iff it minimizes expected objective wrongness according to each of the extremely opinionated politicians that represent the agent’s justificatory state.²⁶

²⁵ Joyce (2010, 288) gives an analogy similar to the one offered here.

²⁶ If supererogation causes problems for SR_{MEOW} (as suggested in the previous section), then it will also cause problems for SR_{MEOW}' . I’ll leave it to the reader to imagine how to modify SR_{MEOW}' in light of those problems (presumably what we’d need is some hybrid of SR_{MEOW}' and SR_{MEOW}^+).

Given SR_{MEOW}' , what shall we say about subjectively *wrong* actions and subjectively *obligatory* actions? I think that a proponent of SR_{MEOW}' should accept the following:

An action A is morally wrong (in the subjective sense) for S(A) to perform if and only if (and because) S(A) can perform A, and A *fails* to minimize expected objective wrongness on each of the justificatory state functions that represent the agent's justificatory state;

An action A is morally obligatory (in the subjective sense) for S(A) to perform if and only if (and because) A *uniquely* minimizes expected objective wrongness on each of the justificatory state functions that represent the agent's justificatory state.

If a proponent of SR_{MEOW}' accepts these additional claims, then she will be committed to saying that when an alternative minimizes expected objective wrongness on *some* but not *all* of the justificatory state functions that represent the agent's justificatory state, the alternative is neither right, nor wrong, nor obligatory (subjectively speaking). She will also be committed to saying the same thing about alternatives that *fail* to minimize expected objective wrongness on some but not all of the agent's justificatory state functions, and for actions that *uniquely* minimize expected objective wrongness on some of the agent's justificatory state functions but fail to minimize it on others. This seems to me to be exactly the right thing to say about actions of these kinds; their subjective moral status seems to me to be indeterminate.

In sum, if our justificatory state turns out not to be precise, as I had previously assumed, this would not be a fatal problem for SR_{MEOW} . The theory could easily be adjusted to accommodate this possibility. To be sure, there are other objections that one undoubtedly could raise to SR_{MEOW} . However, I will not endeavor to address them all

here. For this reason, I do not consider the present chapter to be the final word on the truth of the theory it proposes. I do hope, though, that I have shown that SR_{MEOW} is a promising theory of subjective rightness, one that avoids the problems of Ross's theory as well as the problems of other theories suggested in the literature. At the very least, I hope I have demonstrated that the theory is worthy of continued discussion.

CHAPTER 8

ROSS'S THEORY OF MORAL GOODNESS

8.1. Introduction

In this chapter, I discuss what seems to me to be a neglected topic: W. D. Ross's theory of moral goodness (i.e., his theory of moral worth or moral praiseworthiness). While Ross's theory of moral goodness is of interest for historical reasons, it should also be of interest because of its plausibility. Here, I hope to go at least part of the way toward establishing this. I begin by placing Ross's theory in its historical context—I briefly discuss Kant's theory of moral worth and identify some of its main problems. Then I introduce Ross's theory and argue that it nicely solves Kant's problems. Ross's theory can thus be seen as a plausible response to Kant. In fact, Ross's theory is a *better* response to Kant than others suggested recently in the literature. For instance, Nomy Arpaly and Julia Markovits both suggest that in order to respond to the problems of Kant's theory, we should accept what Markovits calls the "Coincident Reasons Thesis." But this thesis has problems of its own, problems that, I argue, Ross's theory avoids. Ross's theory, then, seems preferable to the Coincident Reasons Thesis. However, Ross's theory is not entirely adequate, at least as it is stated in *R&G*. In *FE*, Ross raises an important problem for his earlier theory, a problem that I call the "nepotism problem." Ross attempts to solve this problem, but I argue that his solution is unconvincing. I then offer a better solution in its place. (In the appendix, I discuss some of the finer details of Markovits's portrayal of the Coincident Reasons Thesis.)

8.2. Ross's Theory (And Some Background)

All of Ross's writings on moral goodness (1928-29, 1930, 1939) include substantial discussion of Kant. Ross seems to have been greatly influenced by Kant, and it is his theory that Ross is chiefly reacting to and trying to improve upon. In fact, Kant's theory of moral worth is the only competing theory that Ross discusses in any detail. For this reason, before turning to Ross's theory of moral goodness, it will be useful to begin by briefly discussing Kant's theory and, especially, its problems. By doing this, the attractions of Ross's theory will shine brighter.

Kant's theory of moral worth (or, at least, *one* component of it) is well-known. Kant famously—or, perhaps, infamously, depending on one's perspective—held that an action has moral worth only if it is done from the motive of duty ([1785] 1997, 3-4 (at 4:390), 11-12 (at 4:398)).¹ While I do not want to get embroiled in the numerous interpretive complexities that surround this proposal, a few clarificatory comments are in order:

1. It is common to distinguish between instrumental and non-instrumental motives.

To illustrate the difference between these two kinds of motives, suppose a building is on fire and some children are trapped inside. Suppose I run into the building with the motive to save the children. However, suppose I'm motivated to save the children only because I'm motivated to receive praise. In this case, my motive to save the children is an instrumental one—I have it only because I'm motivated by something else. However, suppose it's not the case that I'm

¹ It's unclear to me whether Kant would say that this condition is not only necessary but also sufficient for an action to have moral worth.

motivated to receive praise only because I'm motivated by something else.

Suppose I'm motivated for its own sake to receive praise. In that case, my motive (to receive praise) is a non-instrumental one. When Kant said that morally worthy actions must be done from the motive of duty, he surely meant that morally worthy actions must be motivated *non-instrumentally* by the motive of duty.

Henceforth, when I speak of an agent's motives, I will always be referring to her non-instrumental motives.

2. There is debate about whether Kant was committed to the even stronger view that an action has moral worth only if it is done *solely* from the motive of duty.²

However, whether Kant accepted this stronger view will make little difference to what follows.

3. There is also debate about what Kant meant by "the motive of duty." I shall assume that Kant used the phrase to refer to the (true) belief that one's action is morally right. This is a natural interpretation of Kant's use of the phrase, and it's the one that Ross seems to accept (1930, 157-60). However, I suspect that the points I ultimately want to make about Kant's theory will apply on many other interpretations of the phrase as well.³

² See, e.g., Henson (1979) and Herman (1981).

³ For a survey of some of these interpretations, see Stratton-Lake (2000, chaps. 1-4). Note that if Kant held that an action has moral worth only if it is done solely from the motive of duty, and he held that the motive of duty is the belief that the action is right, then he is committed to holding that an action has moral worth only if it is motivated solely by a belief. While Kant may have been comfortable with such a commitment, many philosophers will not be. A long line of philosophers, going back to at least Aristotle, have denied that beliefs alone can motivate. I am

4. Finally, it's not entirely clear what Kant meant by "moral worth" or "moral goodness." I shall use these terms synonymously in this chapter. Furthermore, I will assume that to say that an action has moral worth (or is morally good) is to say that the agent is morally praiseworthy for performing it. Lots of philosophers writing about moral worth follow this practice, though I make no claim here about its historical accuracy, at least with regard to Kant.⁴ Likewise, I shall use the terms "moral badness" and "moral unworth" (as we might call it) synonymously, and I will assume that to say that an action is morally bad (or has moral unworth) is to say that the agent is morally blameworthy for performing it.

There are two important objections to Kant's theory of moral worth. First, Kant's whole-hearted devotion to the motive of duty seems unwarranted; he appears to unjustifiably "fetishize" the motive.⁵ Quite simply, the motive of duty does not seem to be the only motive that can imbue an action with moral worth. To illustrate this point, consider the case of Charitable Charlie. Charlie is walking home late one winter evening, and he sees a homeless woman freezing in the street. Without giving the matter much thought, Charlie gives the woman his coat and hurriedly walks the rest of the way home coatless. Suppose, however, that Charlie didn't give the woman his coat out of the motive of duty: the thought that his action is right never crossed his mind. Rather, he

inclined to agree with these philosophers; it seems to me that both beliefs and desires are needed to motivate action. However, I will largely set this issue aside in this chapter.

⁴ Ross clearly uses the terms "morally good" and "morally worthy" synonymously. He also says that "morally virtuous" means the same as "morally good" (1930, 161) and describes morally good actions as ones that "deserve praise" (32, 45).

⁵ See M. Smith (1994, 75-76).

gave the woman his coat simply because he wanted to alleviate her suffering, and he thought that giving her the coat would do that. Surely Charlie's action has moral worth even though it was not done from the motive of duty. Kant's theory, however, implies that it has no moral worth at all.⁶

Second, Kant's theory has been criticized recently by Julia Markovits (2010, 208) and Nomy Arpaly (2002, 228-31; 2003, 9-11). They argue that Mark Twain's Huckleberry Finn case (1885, 123-28) causes problems for Kant's view. Here's the case they have in mind:

Huck Finn

Huck and Jim are floating on a raft down the Mississippi River. Jim is a slave whom Huck is helping escape, and the two have become close friends on their journey.

However, Huck eventually starts to feel guilty about what he's doing. His conscience asks him, "What had poor Miss Watson [Jim's owner and Huck's caretaker] done to you, that you could see her nigger go off right under your eyes and never say one single word? What did that poor old woman do to you, that you could treat her so mean? Why, she tried to learn you your book, she tried to learn you your manners, she tried to be good to you every way she knowed how. *That's* what she done."

Huck says he "got to feeling so mean and so miserable I most wished I was dead," and he resolves to turn Jim in as soon as possible. This makes him feel "easy, and

⁶ For a related version of this objection, see Stratton-Lake (2000, 11, 51). See also Ross (1939, 306): "And plainly great violence is done to what we really think, when we are asked to believe that ordinary kindness when not dictated by the sense of duty is no better than cruelty. Kant's picture of the ideally good man as going through life never animated by natural kindness but only by the sense of duty has always been felt by most readers to be unduly narrow and rigoristic."

happy, and light as a feather.” However, upon hearing Jim’s excitement about his impending freedom and his sincere expression of gratitude for Huck’s assistance and friendship, Huck starts to feel “sick.” Jim’s kind words, Huck says, take “the tuck all out of me.” Soon Huck has the perfect opportunity to hand Jim over, but Huck finds that he is not “man enough” to do so—he doesn’t have “the spunk of a rabbit.” After squandering the opportunity to turn Jim in, Huck says he feels “bad and low, because I knowed very well I had done wrong.”⁷

According to Markovits and Arpaly (and many others⁸), Huck is praiseworthy for protecting Jim in this case. However, as Markovits and Arpaly point out, Kant’s theory implies that Huck’s action has no moral worth. After all, Huck thinks his action is wrong, not right, and so his action is clearly not motivated by the motive of duty.

Kant’s theory is therefore subject to some important problems. What I’d like to demonstrate now is that Ross’s theory solves these problems. For now, I will focus on Ross’s theory from chapter 7 of *R&G*. Later, I will discuss how he modifies his theory in *FE*.

In its most general form, Ross’s theory of moral goodness in *R&G* can be stated as follows:

⁷ Although this case was authored by Mark Twain, it was brought to the attention of philosophers by Jonathan Bennett (1974).

⁸ See, e.g., Hursthouse (1999, 150-52), Kleist (2009), and Taylor (2012)

RMG: An action is morally good if and only if its agent performs it from a morally good motive.⁹

Ross initially claims that there are three morally good motives: (1) the desire to do what is right, (2) the desire to bring about something good,¹⁰ and (3) the desire to produce pleasure¹¹ or prevent pain in another (1930, 160). However, he goes on to say,

Under (2) I would include the actions in which we are aiming to improve our own character or that of another, without thinking of this as a duty. And believing as I do that a certain state of intellectual nature also is good, I would include actions in which we are aiming at improving our own intellectual condition or that of others. (1930, 160)

So, it seems we should add the following to Ross's list of good motives: (4) the desire to improve someone's character, and (5) the desire to increase someone's knowledge.

At various places, Ross also suggests that love is a good motive (1930, 168-73).

However, it's unclear whether he thinks this is an additional good motive, or whether he thinks it's identical to one already on his list. While Ross doesn't say much about the

⁹ Although Ross never explicitly states RMG in *R&G*, he surely accepts it. What he explicitly says is that "when we ask what is the general nature of morally good actions, it seems quite clear that it is in virtue of the motives that they proceed from that actions are morally good" (1930, 156). However, it becomes clear shortly after Ross says this that he thinks the motives from which good actions spring are good ones (see, e.g., 164-66). Ross is more explicit about this in "The Nature of Morally Good Action," a paper he wrote shortly before writing *R&G*. There he says, "But what is a morally good action? It is, we have said, the doing of something from a good motive" (1928-29, 252).

¹⁰ There is a *de dicto/de re* ambiguity in (1) and (2). Ross surely has in mind the *de dicto* reading.

¹¹ Ross (1930, 166-68) later qualifies this slightly. He says that the desire to bring about a *bad* pleasure in another person is not morally good. (A bad pleasure is, for instance, the pleasure of watching someone suffer.) So, he claims, instead of saying that the desire to bring about pleasure in another is good, what we should say is that the desire to bring about a *non-bad* pleasure in another is good (where a "non-bad" pleasure is a pleasure that is either good or indifferent).

motive of love in *R&G*, he is more revealing in “The Nature of Morally Good Action.” There he takes love to “include any direct interest in the well-being of another person” (1928-29, 254; see also 1939, 115). This seems to be a somewhat different motive from the ones we’ve encountered thus far. So, let’s assume that Ross intends to add it to his list: (6) the desire to improve the well-being of another.

Alongside his theory of moral goodness, Ross appears to accept the following theory of moral badness:

RMB: An action is morally bad iff its agent performs it from a morally bad motive.

Ross claims that there are three morally bad motives: (1) the desire to do what is wrong, (2) the desire to bring about something bad, and (3) the desire to produce pain in another (1930, 166). Although Ross doesn’t discuss it, given what his list of morally good motives ultimately looks like, he’d presumably add the following to his list of morally bad motives: (4) the desire to degrade someone’s character, (5) the desire to decrease someone’s knowledge, and (6) the desire to diminish the well-being of another.¹²

We can now see how Ross’s theory of moral goodness/badness improves on Kant’s theory. First, Ross’s theory does not “fetishize” the motive of duty. On Ross’s view, actions done from the motive of duty do indeed have moral worth. As Ross seems to understand it, the motive of duty includes not only the belief that the action in question is right, but also the desire to act rightly (1930, 157-60), and as we’ve seen, Ross thinks that

¹² In *R&G*, Ross eventually raises a problem for RMB (1930, 166-68). The problem has to do with selfish actions. Selfish actions, Ross thinks, are morally bad, but the motives they spring from (such as the motive to benefit oneself) are not—rather, they are morally indifferent. This prompts Ross to modify RMB slightly. However, I will ignore this complication in what follows.

actions done from this desire have worth. However, Ross thinks that actions done from many other motives (desires) have worth as well. This allows him to say the intuitively correct things about the Charlie case. Recall that Charlie gives his coat to the homeless woman, and he does this not because he wants to do the right thing, but because he wants to alleviate the woman's suffering. According to Ross, this is a morally good motive, and his theory therefore implies that Charlie's action has moral worth. Ross's theory also implies that Huck is praiseworthy for protecting Jim, at least if we assume that Huck protects Jim out of a concern for his well-being (i.e., from the motive of love), from the desire to do something good, or from the desire to please Jim. Ross's theory therefore makes the conclusions about the Huck case that Arpaly and Markovits, and many others, find intuitive.¹³

Ross's theory of moral goodness therefore solves the problems of Kant's theory. However, at this point, I'd like to acknowledge a potential deficiency in Ross's theory. As we've seen, Ross's list of morally good motives is expansive enough to make plausible the claim that Huck and Charlie act from good motives and are thus, according to RMG, doing something that has moral worth. However, one might legitimately wonder whether Ross's list of good motives is complete. Might there be other good motives that Ross overlooks? I think the answer is probably "yes." A few plausible candidates come to mind. Consider the desire to fulfill a promise. Presumably that's a good motive, and an action done from it has moral worth. Or consider the desire to make reparation for a past wrong. That, too, seems like a good motive, and one that can make

¹³ While I agree with Markovits and Arpaly that Huck is praiseworthy for his action, I'll suggest later on that this is only part of the story.

an action morally good. But these motives do not appear on Ross's list¹⁴—according to RMG, then, actions performed solely from them have no moral worth, which may seem difficult to accept.

So, I suspect that Ross's list of morally good motives ultimately needs supplementation. However, I do not think that Ross would be averse to admitting this—in fact, he expands his list of good motives somewhat in *FE* (1939, 296-305), though the list he offers there still seems too anemic, and it doesn't include the desire to keep a promise or the desire to make reparation for a past wrong. However, I will not speculate here on what a complete list of morally good motives might look like. Ross's list seems to me to be at least a good start. And even if his list needs to be made more ecumenical than it currently is, I think he provides us with a plausible general strategy for handling the problems of Kant's theory—namely, claim that the moral goodness/badness of actions is a function of the moral goodness/badness of the motives they are performed on, and claim that there is a plurality of morally good/bad motives.

In fact, Ross's strategy seems like a better strategy for improving on Kant than another that has been suggested recently in the literature. Julia Markovits (2010) and Nomy Arpaly (2002; 2003, chap. 3) both object to Kant's theory and claim that we should accept instead what Markovits calls the "Coincident Reasons Thesis." In the next

¹⁴ That these motives do not already appear on Ross's list is odd. After all, Ross thinks that there are several *prima facie* duties: fidelity (promise keeping), gratitude, reparation, justice, self-improvement, beneficence, and non-maleficence (1930, 21). And, as we've seen, he thinks that alongside the *prima facie* duties of beneficence and non-maleficence, there are corresponding good motives (in particular, the desire to produce pleasure in another and the desire to prevent pain in another, respectively). Why, then, does he not also claim that there are good motives that correspond to the other *prima facie* duties on his list (except, perhaps, for self-improvement)?

section, I will discuss this thesis and point out some of its difficulties. I'll then argue that Ross's theory avoids these difficulties; Ross's solution to Kant's problems is therefore more promising than the one that Markovits and Arpaly propose.

8.3. The Coincident Reasons Thesis

In the introduction of her (2010), Markovits gives the following rough characterization of the Coincident Reasons Thesis: "I suggest an alternative formulation of the thought that morally worthy actions must be performed for the right reasons, according to which morally worthy actions are performed for the reasons *why* they are right" (202). Later, she formulates the thesis more precisely: "According to what I will call the Coincident Reasons Thesis, *my action is morally worthy if and only if my motivating reasons for acting coincide with the reasons morally justifying the action*—that is, if and only if I perform the action I morally ought to perform, for the (normative) reasons why it morally ought to be performed" (205). Arpaly also appears to accept the Coincident Reasons Thesis, though she labels it "Praiseworthiness as Responsiveness to Moral Reasons." She formulates the view as follows: "for an agent to be morally praiseworthy for doing the right thing is for her to have done the right thing for the relevant moral reasons, that is, the reasons making it right" (2002, 226; 2003, 72). Based on these passages, the Coincident Reasons Thesis seems to be the following view:

CRT: An action has moral worth iff the agent performs the action for the reasons why it is morally right (i.e., for the reasons that make it morally right).¹⁵

¹⁵ A similar idea is also present in Stratton-Lake (2000, chap. 1).

The Coincident Reasons Thesis is so called because it claims that morally worthy actions are ones where the agent's motivating reasons¹⁶ *coincide* with the reasons that explain why the action right.¹⁷

Markovits (2010, 208) and Arpaly (2002, 228-31; 2003, 9-11, 75-79) both suggest that the case of Huck Finn causes trouble for Kant's theory of moral worth, and they think that one of the advantages of CRT is that it makes the correct conclusion about the case. At first glance, however, this is puzzling. Huck is supposed to be praiseworthy for protecting Jim. According to CRT, this will be the case only if Huck protects Jim for the reasons why it's right to protect him. But why is it right to help Jim? Of course, that all depends on which moral theory is true. If act utilitarianism is true, for instance, then it's right for Huck to protect Jim because doing so maximizes utility. If Kantianism is true, then it's right for Huck to protect Jim because doing so treats Humanity as an end in itself, not as a mere means. If a (simple) virtue ethics theory is true, then it's right for Huck to protect Jim because a virtuous person would do that in Huck's circumstances. However, Huck's motive for protecting Jim surely isn't to maximize utility, to treat Humanity as an end in itself, or to act as a virtuous person would act. Rather, Huck

¹⁶ As Markovits makes clear (2010, 227-30), the type of motive involved in CRT is the non-instrumental type. In other words, CRT claims that morally worthy actions are ones where the agent's *non-instrumental* motives coincide with the reasons that make the action right.

¹⁷ Markovits and Arpaly think that facts are the sorts of things that make actions right. So, since they think that actions have moral worth only if they are done for the reasons that make them right, they also think that actions are (or at least can be) performed for facts (see Markovits 2010, 221-22). This is a controversial assumption. Some accept this type of view (see, e.g., Dancy 2000). However, others prefer to hold that motivating reasons are psychological states, such as desires, or perhaps complexes of beliefs and desires (see, e.g., M. Smith 1994). Ross seems sympathetic with the psychological state view. However, I will assume that the "fact" view that Markovits and Arpaly presuppose is unproblematic.

protects Jim because he wants to benefit him, or because he wants to help his friend, or something along these lines. So, if act utilitarianism, Kantianism, or virtue ethics is true, then Huck's action has no moral worth according to CRT. The same thing applies if rule utilitarianism, contractualism, divine command theory, or virtually any of the other "main contenders" in normative ethics is true.

Markovits (2010, 226-27) also thinks that CRT is preferable to Kant's theory because it can better handle cases like the Charitable Charlie case. But once again, it's not clear that this really is true. Charlie is presumably praiseworthy for giving the homeless woman his coat when he does so from the motive to alleviate her suffering (and not from the motive of duty). However, according to none of the theories mentioned in the previous paragraph is Charlie's action right in virtue of the fact that it alleviates the woman's suffering. The fact that makes Charlie's action right is, according to these theories, a very different one. So, if any of these theories is true, Charlie's action will have no moral worth on CRT. CRT thus seems guilty of its own type of fetishization. While it doesn't fetishize the motive of duty, it does fetishize whatever fact makes right actions right. According to CRT, only actions done for *that* fact can have moral worth. But this is implausible. It seems that Charlie's action, and others like it, would have moral worth even if utilitarianism, or Kantianism, or virtue ethics, etc. turned out to be true and Charlie's action is therefore not done for the precise reason that explains why it is right.

My main point, so far, is that it seems that in order for CRT to make the correct conclusions about the Huck and Charlie cases, an unusual theory of moral explanation must be correct, one that seems far removed from the theories that have historically

preoccupied normative ethicists. However, there is a natural way for a proponent of CRT to respond to this. She can say that we need to be careful here, for there are different *levels* of moral explanation. She might claim that utilitarians, Kantians, contractualists, etc. are all trying to identify the deepest, most fundamental explanation of rightness; however, in addition to a fundamental explanation of rightness, there is another “shallower” type of explanation. And the rightness of Huck’s action is explained, in this shallow way, by the fact that it helps his friend, by the fact that it improves Jim’s well-being, by the fact that Jim is a person, etc. Likewise for Charlie’s action: the rightness of his action is explained, in the fundamental way, by the fact that it maximizes utility, or by the fact that it treats Humanity as an end in itself, or something like this; however, the rightness of his action is also explained, in the shallow way, by many less complicated facts, such as the fact that it eases someone’s suffering and the fact that it is an act of beneficence. A CRTist might then clarify her view. She might say that when she said that an action has moral worth iff it is performed for the reasons why it is right, she meant to be referring not to the fundamental, strict type of explanation that utilitarians, Kantians, etc. are concerned with, but rather to the non-fundamental, looser type of explanation that is much more permissive. And a CRTist might then claim that once we understand her theory in this way, it makes the correct conclusions about the Huck and Charlie cases. This is because while Huck and Charlie do not perform their actions for the most fundamental reasons why they are right, they do perform their actions for less fundamental reasons why they are right.¹⁸

¹⁸ The response that I outline in this paragraph is similar to the one that Markovits offers. I will discuss the details of her response at length in the appendix (I fear that discussion of those details

While some will undoubtedly take issue with this two-tiered explanatory framework, CRT is problematic even if the framework is justifiable. This is because the rightness of an action cannot be explained, in either the fundamental or the shallow way, by the fact that it is right. There is no type of explanation according to which the fact that an action is right explains why it is right. But this means that if CRT is true, then actions performed solely because they are right have no moral worth. But surely this is implausible. While Kant may have been wrong to hold that the motive of duty is the only motive that can endow an action with moral worth, he was certainly right to suggest that actions done solely from this motive can, at least in some circumstances, have worth. To give one such example, suppose a soldier dives on a live grenade, which kills him but saves five of his compatriots. Suppose the soldier's sole motive is to do the right thing. If that's the case, then CRT implies that the soldier's action has no moral worth. After all, his motive (to act rightly) is not a reason why his action is right. But surely the soldier merits some praise for his action in these circumstances.¹⁹

here would seriously derail the dialectic).

¹⁹ One might object that the case I've described is impossible. Surely, one might claim, it can't be the case that the soldier's *only* motive is to do the right thing. Surely he is also motivated to save his friends, for instance. However, recall that when I speak of a person's motives, I am only talking about her *non-instrumental* motives. So, when I say that the soldier's sole motive is to do the right thing, what I mean is that the soldier's sole non-instrumental motive is to do the right thing. I can admit, however, that the soldier has many instrumental motives, and the motive to save his friends may very well be one of them. However, recall that according to CRT, the moral worth of an action is determined by the agent's non-instrumental motives, not her instrumental ones (see note 16). So, if the soldier's sole non-instrumental motive is to do the right thing, which is the case I am imagining, then CRT implies that his action lacks moral worth. And that, I maintain, is counterintuitive.

Of course, there is more to be said about CRT; the theory is an interesting and plausible one that has been carefully developed and defended by its proponents. However, my goal here is not to decisively refute it. I merely want to make a *prima facie* case for preferring Ross's theory. As we've seen, Ross's theory makes the correct conclusions about the Huck and Charlie cases, and it does so without having to appeal to an unusual or a two-level theory of moral explanation, as CRT does. In addition, the motive of duty is morally good on Ross's theory, so his theory implies that actions performed solely from it have moral worth. It is difficult to see how such actions can have moral worth on CRT. Ross's theory of moral goodness therefore solves the problems of Kant's theory while avoiding the pitfalls of CRT.²⁰ It thus seems, at least *prima facie*, to be a more attractive alternative to Kant's theory than CRT is.

8.4. The Nepotism Problem

While Ross's theory of moral worth is a plausible alternative to Kant's theory, it is not without its own problems. I'd now like to focus on one of these problems, one that Ross himself identifies in his later work. In contrast to what he holds in *R&G*, in *FE*, Ross says that "to say that the goodness of actions depends *solely* on the goodness of their motives would be to simplify matters far too much" (1939, 306-7). To demonstrate this, Ross asks us to consider the following hypothetical case. Suppose a person (A) knows that an action of his will bring about pleasure for one person (B) but will bring about substantially more pain for two other people (C and D). Suppose, however, that A performs the action purely from the motive to bring B pleasure. According to Ross's

²⁰ I raise an additional problem for CRT in note 26.

theory of moral goodness from *R&G*, A's action is morally good since it springs from a morally good motive. But, Ross says, this implication is counterintuitive. In light of this, he proposes a new theory of moral goodness. He says:

The way in which we judge of the goodness of an action is, I think, somewhat as follows. If *A* does an act which he foresees to be likely to have certain characteristics, we ask ourselves what attractions an ideally good man would have towards the act in virtue of certain of its characteristics, and what aversions he would have in virtue of others. We judge, perhaps, that an ideally good man would be more deterred from the act because it would hurt *C* and *D* than he would be attracted towards it because it would give pleasure to *B*; and we judge *A*'s action bad on the whole not because of its actual motive, which is good, but because in doing it *A* is failing to have a strong aversion which an ideally good man would have. We judge the action by comparing the agent's set of attractions and aversions with the set of attractions and aversions which would ideally arise in face of the foreseen changes to be produced by the action. (1939, 307)

To further illustrate this problem and his intended solution to it, Ross gives another example. Suppose that person *A* is in charge of hiring someone for a job, and persons *B*, *C*, and *D* are among the candidates. *C* and *D* are the best candidates for the job, and *A* realizes this. However, *A* ends up hiring *B* for nepotistic reasons. According to Ross's theory of moral goodness from *R&G*, *A*'s action is morally good since he does it from a good motive (namely, to benefit *B*). But, Ross says, "his [*A*'s] action is definitely bad, because he is not being deterred as an ideally good man would be by the thought of the injustice to *C*, *D*, and the rest" (1939, 307). (I shall refer to the problem for RMG/RMB that Ross raises here as the "nepotism problem." It will also be convenient to give person *A* a name. I will call him "Ned the Nepotist," or just "Ned" for short.)

The theory of moral goodness that Ross suggests in these passages can perhaps be stated as follows:

An action is morally good iff its agent performs it from the motives that an ideally morally good person would have in the agent's circumstances.

This, however, is subject to an important qualification. Ross later acknowledges that if an action were performed from the motives that an ideally good agent would have in the circumstances, the action would be *ideally* (completely, perfectly) good (1939, 309). The above theory, then, should be construed as a theory of the *perfect* moral goodness of actions. However, Ross also thinks that actions can be partially morally good. Presumably, he'd say that an action is partially good iff it is performed from the motives that a partially good agent would have in the circumstances. More precisely:

RMG_{FE}: An action is morally good to degree X iff its agent performs it from the motives that a morally good person of degree X would have in the agent's circumstances.

Although Ross doesn't discuss it, he would presumably accept the following corresponding theory of moral badness:

RMB_{FE}: An action is morally bad to degree X iff its agent performs it from the motives that a morally bad person of degree X would have in the agent's circumstances.

RMG_{FE} and RMB_{FE} are significant departures from the views Ross held in *R&G*. In *R&G*, Ross notes that many things seem to be morally good: people, characters, actions, and feelings (1930, 155). However, Ross focuses his attention on morally good actions; he thinks that once we have an account of morally good actions, we will be able to use it to construct plausible accounts of the moral goodness of other things, such as persons, characters, and feelings. In *R&G*, Ross therefore takes the moral goodness of actions to

be explanatorily prior to the moral goodness of these other things. But this is no longer the case in *FE*. As we've seen, in that work, Ross understands the moral goodness of actions in terms of the moral goodness of persons. The moral goodness of persons therefore now precedes, explanatorily, the moral goodness of actions.

What shall we think of this shift in position? Are RMG_{FE} and RMB_{FE} plausible? There's reason to believe that the answer is "no."²¹ According to these theories, an action is morally good/bad to a certain degree iff it is performed from the motives that a person who is good/bad to that degree would have in the agent's circumstances. But it's misleading to talk about *the* motives that a morally good/bad person would have in a situation. Different morally good people might have different motives in any given situation—similarly for morally bad people. Return to the nepotism case: for nepotistic reasons, Ned gives B a job even though Ned recognizes that C and D are better candidates. What motives would a person with a high degree of moral goodness have in this situation? It's difficult to say, especially given the fact that morally good people may, at least on occasion, have *bad* motives. After all, even the world's pillars of moral goodness—the Nelson Mandelas, the Mother Teresas, the Gandhis—don't live lives entirely without blemish. It's thus possible for a very good person to have a nepotistic motive and to act on it, from time to time. A very morally good person might therefore act just as Ned does, and from exactly the same motives. Of course, a morally good

²¹ Hurka (2006) argues that it's more plausible to understand the moral goodness of persons (or dispositions) in terms of the moral goodness of actions, instead of the other way around. He would therefore be unhappy with Ross's shift in position, especially since he seems sympathetic to Ross's philosophical outlook. The objection I'm about to raise to RMG_{FE}/RMB_{FE} is inspired by Hurka's paper.

person might also act differently and have different motives. But my point is that this is not a foregone conclusion. Different morally good people, even those who are very morally good, might have different motives in Ned's circumstances, and some of these people will undoubtedly have motives that are similar to the ones that Ned actually has. For this reason, it's not clear what RMG_{FE} implies about the moral goodness of Ned's behavior.

It's likewise unclear what RMB_{FE} implies about the moral badness of Ned's action. Just as morally good people will occasionally have bad motives, morally bad people will occasionally have good motives. After all, even the evil demons of the world—the Hitlers, the Stalins, the Dahmers—surely don't live lives of complete stain. So, in any given situation, morally bad people, even extremely bad people, will have a variety of motives, many of which are bad, but some of which are good. It is therefore unclear whether RMB_{FE} implies that Ned's nepotistic act is in fact morally bad, as it intuitively seems.

Perhaps there is a way of modifying RMG_{FE} so it avoids these problems.²²

However, instead of dwelling on the view any further, I'd like to look at another, *prima*

²² Consider, for instance, the following variant of RMG_{FE} / RMB_{FE} :

RMG_{FE}' : An action is morally good to degree X iff its agent performs it from motives that approximate to degree X the motives that a completely morally good person would have in the circumstances.

RMB_{FE}' : An action is morally bad to degree X iff its agent performs it from motives that approximate to degree X the motives that a completely morally bad person would have in the circumstances.

Arguably, these variants are an improvement because while different morally good (bad) people might have different motives, each completely morally good (bad) person has the same motives.

facie more plausible, way of solving the nepotism problem. In particular, it may appear that we can solve the problem by building rightness into Ross's theory of moral goodness from *R&G*. Consider, for instance, the following view:

RMG+: An action is morally good iff the action is morally right, and its agent performs it from a morally good motive.²³

RMG+ implies that Ned is not praiseworthy for giving the job to his relative since although he has a good motive, his action is presumably not right. RMG+ also makes the desired conclusions about the other cases mentioned so far. It implies that Huck is praiseworthy for helping Jim and Charlie is praiseworthy for giving away his coat since these actions are right and done from good motives.

Of course, a proponent of these theories will need to say more about what it is for a person's motives to "approximate" the motives of a completely morally good (bad) person, which may be a difficult task. In addition, it's not clear that these theories yield the desired implications about the nepotism case. Ned, by stipulation, has only good motives; his sole motive in acting is to benefit his relative. This motive, according to Ross, is a good motive, though it's not the best one. The highest motive, according to Ross, is the motive of duty, and Ross would presumably say that a completely morally good person would always act from it. Still, it seems that Ross would have to admit that Ned's motives approximate, to some degree, the motives of a completely morally good person. RMG_{FE}' thus implies that Ned's action is morally good to some degree. But that's not the result we're after; we're trying to explain why Ned's action is morally *bad*, not morally good. Ross might argue that Ned's motives also approximate, to some degree, the motives of a completely morally bad person, and so Ned's action is, according to RMG_{FE}'/RMB_{FE}', morally good *and* morally bad. As we'll later see, I'm sympathetic to this idea. The problem is that RMG_{FE}'/RMB_{FE}' appears to imply the same thing about Huck's action. Indeed, since Ned and Huck have, by stipulation, the same motives, the theory seems to imply that their actions have the same degree of moral goodness and the same degree of moral badness. But I'm much less sympathetic to that idea.

²³ Presumably, a proponent of RMG+ will say that the degree to which an action is morally good is a function of the degree to which its motive is morally good.

RMG+ only gives us conditions for the moral goodness of actions. What would a corresponding theory of moral badness look like? At first glance, one might think it should look like this:

An action is morally bad iff the action is morally wrong, and its agent performs it from a morally bad motive.

While this may be the most natural way of formulating a theory of moral badness that corresponds to RMG+, it does not withstand scrutiny. In the nepotism case, it implies, counterintuitively, that Ned's action is not morally bad since although his action is wrong, he doesn't, by stipulation, perform it from a bad motive. For this reason, the following theory is a better counterpart to RMG+:

RMB+: An action is morally bad iff the action is morally wrong, *or* its agent performs it from a morally bad motive.²⁴

RMB+ makes the correct conclusion about the nepotism case, as should be clear, and it also makes the correct conclusions about the Huck and Charlie cases. Both Huck and Charlie perform right actions from good motives; RMB+ therefore implies that their actions are not morally bad.

RMG+ and RMB+ therefore seem to be plausible ways of developing Ross's theory of moral goodness and badness in light of the nepotism problem. However, they have an important problem. On RMG+/RMB+, blameless wrongdoing is not possible. After all, according to RMB+, acting wrongly is sufficient for one's action to be morally bad. But

²⁴ Presumably, a proponent of RMB+ will say that the degree to which an action is morally bad is a function of the degree to which it is morally wrong and/or the degree to which its motive is morally bad.

I've been assuming that to say that an action is morally bad is to say that its agent is blameworthy for performing it. So, if RMB+ is true, every wrong action will be one for which its agent is blameworthy. However, many philosophers have argued—including Ross, at least in his early work (1930, 7, 45, 156)—that it's possible for an agent to be blameless (or even praiseworthy) for doing the wrong thing.²⁵ Here's a typical example:

Brave Brenda

On her walk home from work, Brave Brenda witnesses a terrible car accident. She tries to call 911 but has no cell phone service, and no one else is around to alert. One of the wrecked cars is on fire and it looks to Brenda as though its inhabitants will not survive long. So, at great risk to her own safety, Brenda rushes to the car and pulls one of the victims from the burning wreckage. Suppose, however, that pulling the victim to safety only further aggravates the spinal injury he sustained during the crash, paralyzing him for life. Had Brenda left the victim where he was, he would have been much better off—the fire subsided much faster than Brenda had anticipated, and the other people in the car were, for the most part, fine.

In this case, it seems that Brenda did the wrong thing—she shouldn't have pulled the victim from the car. However, Brenda is surely blameless for her action (indeed, she seems praiseworthy for it). The case, then, appears to be a case of blameless

²⁵ See also Moore ([1912] 2005, 100-1), Zimmerman (2004), Haji (1997, 528-29; 1998, 146), and Graham (2010, 94).

wrongdoing. As I noted, if there can be cases of this type, then RMB+ is problematic since it implies that these cases are impossible.²⁶

However, there is a natural way for a proponent of RMG+/RMB+ to respond to this, at least if she accepts the distinction between objective and subjective rightness/wrongness that Ross makes. If she accepts this distinction, then she can point out that her theory, as it is currently stated, is unclear. This is because the terms “right” and “wrong” in her theory are ambiguous—they have both an objective and a subjective sense. And, she can go on to say that she meant these terms to express their *subjective* senses in her theory, not their objective senses. Given that, a proponent of RMG+/RMB+ can argue that the Brave Brenda case, and other cases like it, poses no problem for her theory. This is because although Brenda’s action is objectively wrong (wrong in light of the facts of Brenda’s case), it is subjectively right (right in light of Brenda’s beliefs or evidence about her situation). A proponent of RMG+/RMB+ can say that in general, blameless wrongdoing is possible when the sense of “wrongdoing” at issue is the objective sense. However, blameless wrongdoing is not possible when the sense of “wrongdoing” at issue is the subjective sense. The case of Brave Brenda is not a counterexample to this, nor are any of the other paradigmatic cases of blameless wrongdoing. The paradigmatic cases of blameless wrongdoing are really only examples of blameless *objective* wrongdoing; they are not examples of blameless *subjective* wrongdoing.

²⁶ Notice that this case also causes problems for CRT. Since Brenda’s action isn’t right, Brenda does not act for the reasons why it’s right. CRT therefore implies that Brenda’s action lacks moral worth.

But if a proponent of RMG+/RMB+ responds in this way to the Brenda case and other purported cases of blameless wrongdoing, she will face a different problem. This is because if the sense of “right” and “wrong” at issue in RMG+/RMB+ is the subjective sense, then it’s no longer clear that the theory will be able to handle the Huck case. According to this construal of RMG+/RMB+, Huck’s action is morally good only if it is subjectively right (that is, only if it is right in light of his mental state, in particular his beliefs and evidence). But it’s hard to believe that his action is in fact subjectively right. After all, he thinks his action is wrong. He probably has good reason to think his action is wrong too, given the society he lives in and the moral role models he has (e.g., Miss Watson (who owns slaves), Tom Sawyer, Pap, etc.). For this reason, it seems to me that we should say that Huck’s action is subjective wrong, not right. And if that’s the case, then the current understanding of RMG+/RMB+ implies that Huck’s action is morally bad, not good.

It’s worth noting here that Ross’s own theory of subjective rightness appears to imply that Huck’s action is not subjectively right. As we saw in chapter 4, Ross’s view is that an action is subjectively right iff its agent believes it maximizes the net balance of prima facie rightness over wrongness. But it’s very hard to believe that Huck thought his action maximizes this balance. Insofar as Huck had any beliefs about this balance, surely he thought that turning Jim in, rather than helping him, maximizes it.

Other plausible theories of subjective rightness also seem to imply that Huck’s action is not subjectively right. Consider, for instance, the following consequentialist theory of subjective rightness: an action is subjectively right iff its agent believes it maximizes the net balance of goodness over badness. Of course, we don’t have access to

the content of Huck's mind, but it is difficult to believe that Huck thought his action maximized the net balance of goodness over badness. After all, Jim tells Huck that once he is free, he will save his money to buy his children, and if their master won't sell them, he will "get an Ab'litionist to go and steal them" (Twain 1885, 124). Huck says it "froze me to hear such talk" and that it "was such a lowering of him [Jim]" (124)—Huck seems to regard the theft of children out of slavery as a terrible thing. In part because of this, it seems to me much more likely that Huck believes that turning Jim in will have better overall consequences than will helping him escape, rather than the other way around.

Consider now the following virtue ethics theory of subjective rightness: an action is subjectively right iff its agent believes that a virtuous person would perform the action in the agent's circumstances. But Huck surely does not believe that a virtuous person would, in Huck's situation, help Jim escape. If anything, he believes that a virtuous person would turn Jim in.

Or consider a Kantian theory: an action is subjectively right iff its agent believes the action's maxim can be consistently willed to be a universal law of nature. I think it's a stretch to attribute to Huck any belief about whether the maxims of his actions can be consistently willed to be a universal law of nature. And even if we can attribute such a belief to him, it's not at all clear to me that we should say that he thought that the maxim associated with protecting Jim can indeed be consistently willed to be a universal law.

That's enough theories to make my point, which is this. I've said that when we look at Huck's beliefs and evidence, it's difficult to maintain that it's subjectively right for him to help Jim. And this difficulty is underscored when we start to look at plausible theories of subjective rightness. We've seen that when we look at natural ways of

articulating a Rossian, a consequentialist, a Kantian, or a virtue ethics theory of subjective rightness, Huck's action does not appear to be subjectively right. Of course, there are other plausible theories to consider. Perhaps one of them will imply that Huck's action is subjectively right. But I think the burden is on the proponent of RMG+/RMB+, or on anyone else who wants to say that Huck's action is subjectively right, to convince us otherwise.

8.5. The Solution

We're starting to see that the Huck case, the nepotism case, and the Brenda case together cause a problem for the Rossian. Huck's and Brenda's actions seem morally good; Ned's action seems morally bad. However, it's difficult to give a Rossian theory that has each of these implications. RMG/RMB makes the correct conclusions about the Huck and Brenda cases, but it founders on the nepotism case. The objective version of RMG+/RMB+ makes the correct conclusions about the Huck case and the nepotism case, but it has trouble with the Brenda case. On the other hand, the subjective version of RMG+/RMB+ yields the intuitively correct implications about the nepotism case and the Brenda case, but not the Huck case. RMG_{FE}/RMB_{FE} perhaps makes the correct conclusions about each case, but it suffers from other serious problems.

In this section, I want to sketch a way forward. What I want to suggest is that Ross's theory from *R&G* is, by and large, correct. However, it captures only one part of the story about moral goodness/badness. There are different kinds, or types, or dimensions of moral goodness/badness, and Ross's theory from *R&G* accounts for only one of them.

In order to flesh out this idea, let me first say a bit more than I have already about how I'm understanding moral goodness and badness. As I mentioned in sec. 8.2, when I say that an action is morally good (or has moral worth), what I mean is that the agent is morally praiseworthy for performing it. Likewise, when I say that an action is morally bad (or has moral unworth), what I mean is that the agent is morally blameworthy for performing it. So, then, I'm ultimately understanding moral goodness and badness in terms of moral praiseworthiness and blameworthiness, respectively. I think this is a plausible way of interpreting Ross's (and Kant's) use of the terms "moral goodness," "moral worth," and the like.

However, what exactly do I mean when I say that a person is "praiseworthy" or "blameworthy" for doing something? Well, when I say that a person is praiseworthy for performing an action, I mean that the person is deserving of praise for performing it, and when I say that a person is blameworthy for performing an action, I mean that the person is deserving of blame for performing it. Furthermore, when I use the terms "praise" and "blame" in this context, I am, in each case, referring to a moral reactive attitude. There will undoubtedly be debate about just how to understand the reactive attitudes to which "praise" and "blame" refer, but for my purposes, I needn't dwell on the matter. It will suffice for me to say that "praise" refers to a positive reactive attitude such as moral approbation or commendation, and "blame" refers to a negative reactive attitude such as moral resentment or indignation. Thus on my view, to be praiseworthy for performing an action is to be deserving of moral approbation (or something similar) for performing it, and to be blameworthy for performing an action is to be deserving of moral resentment (or something similar) for performing it. This view is in no way an idiosyncratic one—it

is a view firmly within the “Strawsonian” tradition, a tradition that began with P. F. Strawson’s landmark paper, “Freedom and Resentment” (1962), and has since been developed by many philosophers, including, most notably, R. Jay Wallace (1994).

Since my account of praiseworthiness/blameworthiness makes use of the concept of desert, it behooves me to say a few words about the concept. Unfortunately, I don’t have much to say about the concept of desert, for it seems to me to be a primitive one, or at least close to it. However, I think we can say a few general things about it. First, it is standardly assumed that the desert relation is a three-place relation that has as its relata a subject, an object, and a desert base. So, for instance, I (the subject) might deserve an apology (the object) because I have been wronged (the desert base).²⁷ This standard assumption, however, has been challenged.²⁸ The problem with the assumption is that it is natural to say things that suggest that the desert relation has additional places. For instance, it would be natural to say that I deserve an apology *from you* because you’ve wronged me. So maybe we should hold that the desert relation has four places: a subject, two objects, and a desert base. But notice that it is also natural to mention time indexes when making desert claims. We might say that I deserve an apology from you *now* because you wronged me *yesterday*. So, perhaps we should hold that the desert relation has six places: a subject, two objects, two time indexes, and a desert base. On the other

²⁷ As Serena Olsaretti says, “A glance at a few of the most important contemporary contributions on desert and justice reveals a remarkable variety of different approaches to the topic, with only a few basic points of substantive agreement among desert theorists. These agree that desert is a three-place relation between a person, the grounds on which she is said to be deserving (the desert basis), and the treatment or good which she is said to deserve (the deserved good)” (2003, 4; see also McLeod 2013).

²⁸ See Feldman (in progress).

hand, when we make a desert claim it's also common to omit some of these things. We might simply say that I deserve an apology from you now. So maybe we should hold that the desert relation has fewer places after all. For my purposes, however, it doesn't matter how many places the desert relation has. All I want to say is that whenever something is deserved, there is a thing that deserves it (namely, a subject); whenever someone is deserving, there is something that the person deserves (namely, an object); and whenever someone deserves something, there is a consideration in virtue of which the person deserves it (namely, a desert base). It makes no difference to me whether these things are places in the desert relation, or whether the desert relation has additional places as well.

Let me now say a few things about desert bases. First, desert bases are plausibly facts. Suppose, for instance, that I deserve to get the job. I might deserve to get it because I am exceptionally qualified for it. In this case, the fact that I am exceptionally qualified for the job is the desert base upon which I deserve to get the job. Or suppose that I deserve to receive the gold medal in the 100-meter backstroke. I might deserve to receive it because I won the 100-meter backstroke race. In this case, the fact that I won the race is the desert base upon which I deserve to receive the medal.

An interesting feature about desert bases is that it can sometimes happen that there are several bases in virtue of which a person deserves something. Suppose, for instance, that I deserve to receive health insurance. There might be multiple bases upon which I deserve this. Perhaps I work for a company that has promised to provide its employees with health coverage. In that case, it seems that I deserve to receive health insurance in virtue of my working for the company in question. But, one might think that I also deserve health care even if I do not work for a company that provides it to its employees.

One might think that health insurance is close to a basic human right—anyone who belongs to a state deserves to receive health insurance from the government. If that's the case, then I might also deserve to receive health insurance simply in virtue of my belonging to a state. Consider another, similar example, one that's closer to Huck's world. Suppose a slave in Huck's time works hard for many years, scrupulously saves his money, and is eventually able to buy his own freedom. In that case, it seems that the slave deserves his freedom in virtue of his purchase of it. But, of course, the slave deserves to be free even if he does not buy it: plausibly, the slave deserves his freedom simply in virtue of his being human.

That there can be multiple bases upon which a person might deserve something is important for my purposes because I'd like to suggest that this is precisely what happens with praise and blame. That is, there are several considerations in virtue of which a person might deserve praise, and there are several considerations in virtue of which a person might deserve blame. Praise and blame have multiple desert bases. Moreover, what I want to suggest is that Ross was right about one of these desert bases. That is, one of these desert bases has to do with a person's motives: a person deserves blame for performing an action when she performs it from a bad motive; a person deserves praise for performing an action when she performs it from a good motive. Another of these desert bases, I propose, has to do with a person's beliefs about the deontic status of her action: a person deserves blame for performing an action when she performs it in the

belief that it is wrong; a person deserves praise for performing an action when she performs it in the belief that it is obligatory.²⁹

If there can be multiple desert bases for praise and blame, and if these desert bases are as I've suggested, then this provides us with a potential solution to the problem that plagued the Rossian. If one desert base for praise has to do with one's motives, then I can grant that Huck and Brenda are praiseworthy for their actions (in virtue of their performance of their actions from good motives). On the other hand, if a desert base for blame has to do with one's moral beliefs, then I can also grant that Ned is blameworthy for his action (in virtue of his belief that he's acting wrongly—surely Ned has such a belief; if he lacks it, then it no longer seems at all clear that he's blameworthy). So my proposal has the desired result: it allows us to say that Huck and Brenda are praiseworthy for their actions, but Ned is blameworthy for his.

Of course, while my proposal allows us to say these things, it also forces us to say *more*. It forces us to admit that while Huck deserves praise for his action, he also deserves blame for it (in virtue of his performance of it in the belief that it is wrong). And it forces us to admit that while Ned deserves blame for his action, he also deserves praise for it (in virtue of his performance of it from a good motive). (On my proposal, Brenda is *doubly* praiseworthy—she deserves praise in virtue of her motives and her beliefs.)

²⁹ Some might doubt that anyone deserves praise just for doing something she thinks is obligatory. For instance, if a mother takes care of her child thinking that she ought to, does she deserve praise for her action in virtue of this fact? Maybe not. Perhaps, then, what we should say is that a person deserves praise for performing an action when she performs it in the belief that it is *supererogatory*. Thanks to Pete Graham for pressing me on this point.

Should it bother us to say these additional things? With regard to the Huck case, I think the answer is clearly “no.” In fact, I think it’s a good thing that we can say that Huck is both praiseworthy and blameworthy for his action. When I think about the Huck case, my intuitions pull me in different directions. Part of me wants to commend Huck. After all, he has chosen to help a friend even though doing so is dangerous. But part of me also wants to criticize Huck. After all, he’s flagrantly disobeying his conscience. My proposal accommodates this tension nicely, for according to it, Huck is in fact commendable for his action (he’s praiseworthy for it in virtue of his motives), but he’s also criticizable for his action (he’s blameworthy for it in virtue of his beliefs). So, the fact that my proposal forces us to say that Huck is both praiseworthy and blameworthy for his action should be welcomed, not resisted—at least on my view. The proposal accounts nicely for the tension that I feel when I consider the case.

Turning now to the nepotism case, is it a problem to say that Ned is both praiseworthy and blameworthy for his action? Here things are a bit more complicated because I don’t feel as much tension when I think about the case. My intuitions are less divided. I definitely want to criticize Ned, but I feel less pressure to commend him too. However, on reflection, I think we should admit that if Ned really does act from good motives—if he’s really just trying to help out his relative—then he does in fact deserve some praise for what he’s done. However, I also want to say that the amount of praise he deserves here is small. He deserves much *more* blame than praise. The degree to which he’s blameworthy for his action *overshadows* the degree to which he’s praiseworthy for it. Perhaps that’s why it’s easy to overlook the fact that he’s praiseworthy. Since he is much more blameworthy than praiseworthy, it’s easy to ignore his praiseworthiness.

But if we are going to say that Ned is more blameworthy for his action than he is praiseworthy for it, then it might seem that we will have to say the same thing about Huck. After all, both Ned and Huck believe that their actions are wrong. So, the degree to which their actions are blameworthy in virtue of being performed in spite of this belief is presumably the same (or nearly so). Ned and Huck also have very similar motives. So, the degree to which their actions are praiseworthy in virtue of their motives is also presumably the same (or nearly so). So, if we're going to maintain that the blameworthiness of Ned's action swamps its praiseworthiness, then don't we also have to maintain (rather implausibly) the same thing about Huck's action?

Well, yes, unless we can find a relevant difference between the cases. But I think there is an important difference between them. The difference has to do with the virtues that Huck and Ned exhibit when they act. One virtue that is relevant in both cases is the virtue of loyalty. Both protagonists exhibit loyalty to their nearest and dearest when they act. But, Huck exhibits substantially more loyalty than Ned does. To see this, I'd like consider another memorable passage from the story, one that appears after the passage I cited earlier. In the passage I have in mind (Twain 1885, 269-72), Huck is again feeling guilty about helping Jim. He says that if anyone finds out that he has helped a slave escape, Huck will have to "be ready to get down and lick his boots for shame." Moreover, he says that something inside him keeps telling him that he could have gone to Sunday school, and if he did, he would have learned full well "that people that acts as I'd been acting about that nigger goes to everlasting fire." So, Huck resolves to write a letter to Miss Watson to let her know where Jim is. But after he writes the letter, Huck reflects on his adventures with Jim and all of the nice things that Jim has done for him, and he

starts to wonder whether he should send it. After wrestling with the matter for a while, Huck says, “I took it [the letter] up, and held it in my hand. I was a trembling, because I’d got to decide, forever, betwixt two things, and I knowed it. I studied a minute, sort of holding my breath, and then says to myself: ‘All right, then, I’ll *go* to hell’—and I tore it up. It was awful thoughts, and awful words, but they was said. And I let them stay said; and never thought no more about reforming.” So, Huck is not only willing to risk public shame for Jim, but he’s also willing to risk going to Hell for him. Huck therefore exhibits extraordinary loyalty to his friend. This loyalty is much deeper than the loyalty that Ned exhibits when he chooses to hire his relative over the more qualified candidate. Huck’s loyalty is deeper because he risks more. Ned risks losing his job perhaps. But Huck risks his eternal salvation.

Earlier I claimed that there are two bases upon which a person can deserve praise or blame: one has to do with the person’s motives, the other has to do with the person’s beliefs. But I’d now like to suggest that there can be other bases as well. In particular, I think that a person deserves praise for an action when she exhibits a virtue while performing it. Likewise, a person deserves blame for an action when she exhibits a vice while performing it. So, since both Ned and Huck exhibit the virtue of loyalty when they act, they both deserve praise for their actions because of this. But since the amount of loyalty that Huck exhibits is greater than the amount of loyalty that Ned exhibits, Huck deserves much more praise for his action on this score than Ned does. All told, then, while Huck and Ned both deserve praise and blame for their actions, Huck, I maintain, deserves more praise for his action than Ned does because of Huck’s greater loyalty.

Are other virtues relevant in these cases? Perhaps. Huck may exhibit a certain type of moral courage that Ned doesn't exhibit, though I'm not entirely sure about this. What makes me hesitate to attribute any type of courage to Huck is his belief that he's acting wrongly. Risking your eternal salvation to help someone escape from slavery may sound like an extraordinarily courageous thing to do, but it seems significantly less courageous if you think, as Huck does, that you're acting wrongly in the process. In fact, Huck seems to exhibit a type of moral cowardice—he's unwilling to do what he thinks he ought to do. (Likewise for Ned.)

But the matter is complex because it's not clear how exactly to understand Huck's psychology. It seems pretty clear from the text that Huck has a conscious belief that it's wrong to help Jim escape. But I think a case can be made that Huck also has an unconscious belief, or perhaps an unconscious inclination to believe, that he ought to help Jim. If Huck believes, consciously, that he should turn Jim in, but he also believes, deep down, that he should continue helping Jim escape, this would help explain why Huck struggles so much to make his decision. The thought here would be that Huck has difficulty deciding whether to turn Jim in because his conscious beliefs are colliding with his unconscious beliefs (or, perhaps, his unconscious dispositions). Hursthouse (1999, 150-53) interprets the Huck Finn case in this manner.

But if Huck does have an unconscious recognition that he ought to help Jim, and Huck chooses to help Jim even though he thinks it will land him in Hell, then it *does* seem that Huck is displaying a certain kind of moral courage. But notice that nothing like this is plausibly going on in the Ned case. It's not as though Ned believes, deep down, that he ought to hire his relative. Rather, Ned knows it's wrong to hire his relative

over the better qualified candidate, but he does so anyway. No moral courage is plausibly displayed here. But if Huck does in fact display courage when he acts, then this gives us another way to differentiate the Huck case from the Ned case: Huck deserves yet more praise for his action in virtue of the moral courage he displays; not so for Ned.

To summarize, it will be useful to look at Huck and Ned side by side. I've suggested that both Huck and Ned deserve blame for their actions in virtue of their belief in wrongdoing, and they both deserve praise for their actions in virtue of their good motives. However, I've also suggested that the exhibition of virtue/vice while acting is a further basis upon which one can deserve praise/blame for one's action and that Huck and Ned therefore both deserve additional praise for their actions in virtue of their exhibition of loyalty to their nearest and dearest. However, I've argued that Huck deserves more praise for his action in virtue of his exhibition of loyalty since his loyalty is deeper. Additionally, I've speculated that Huck might deserve more praise still if he can properly be said to display moral courage when he acts. All told, then, Huck deserves more praise for his action than Ned does for his, and so even if Ned ultimately deserves more blame than praise for his action, as I've suggested, we don't necessarily have to say the same thing about Huck.

8.6. An Objection

Let me end by considering a general objection to the view I am proposing. I've maintained that there are multiple bases upon which a person can deserve praise or blame. This implies that a person can be both praiseworthy and blameworthy for one and the same action—indeed, that's exactly what I've claimed happens in the Huck and Ned

cases. But one might object that this is impossible: no one can be both praiseworthy and blameworthy for doing one and the same thing.

But why should we think that this is impossible? I can think of three reasons. First, one might argue that it's impossible for someone to be both praiseworthy and blameworthy for an action because to say that a person is blameworthy for an action just is to say that the person is not praiseworthy for it. So, if a person can be both praiseworthy and blameworthy for an action, that means that a person can be both praiseworthy and not praiseworthy for it, which is impossible.

But this line of reasoning is flawed. "Blameworthy" doesn't mean "not praiseworthy." Someone can fail to be praiseworthy for doing something without being blameworthy for it. For instance, when I ate my oatmeal this morning, I failed to be praiseworthy for my action. But neither was I blameworthy for it. Most of our actions are like this, I think—that is, most of our actions are ones that we are neither praiseworthy nor blameworthy for. In any case, I have carefully articulated how I am understanding the terms "praiseworthy" and "blameworthy" in this chapter. To repeat, for a person to be praiseworthy for doing something is for the person to deserve praise for it; for a person to be blameworthy for doing something is for the person to deserve blame for it. I see why a person can't be both praiseworthy and not praiseworthy for performing an action. I also see why a person can't deserve praise and not deserve praise for an action. But I don't yet see any good reason for thinking that a person cannot deserve both praise and blame for an action and thus, on my definition, be both praiseworthy and blameworthy for it.

Second, one might argue that it's impossible for a person to be both praiseworthy and blameworthy for one and the same action because it's psychologically impossible to simultaneously praise and blame a person for one and the same action. If it's psychologically impossible to simultaneously praise and blame someone for an action, then, the thought goes, a person can't simultaneously deserve praise and blame for it, and she therefore can't be simultaneously praiseworthy and blameworthy for it.

But this is a bad argument. Someone can *deserve* something even if she can't be given it. A child, for instance, might deserve a parent's love even if the parent is incapable of loving the child. A criminal might deserve to be punished even if we can't punish him (because, say, he's gone into hiding or has fled the country). Thus, I think a person can deserve something even if it's not possible to give it to her, and so even if it's true that we can't both praise and blame someone for an action, it doesn't follow that the person doesn't deserve both praise and blame for it.

Finally, one might argue that my suggestion that a person can be both praiseworthy and blameworthy for an action is impossible because praiseworthiness implies right-doing and blameworthiness implies wrongdoing. My suggestion therefore implies that an action can be both right and wrong at the same time. But, one might think, this is impossible.

However, we need to be careful here. What sense of "right" and "wrong" is being used here, the objective sense or the subjective sense? If the objective sense is at issue, then the proposed link between praiseworthiness/blameworthiness and rightness/wrongness is implausible. As the Brenda case demonstrates, a person can be praiseworthy for performing an objectively wrong action. There are also clear cases of

blameworthy right-doing. Here's a standard case of this type (based on Graham 2010, 94).³⁰ Suppose a soldier has a gangrenous leg that will soon kill him if left untreated. The soldier begs the military surgeon to amputate the leg even though the surgeon has no anesthetics. The surgeon agrees and gleefully saws off the soldier's leg. However, the surgeon does not amputate the soldier's leg because he wants to save the soldier's life; rather, he amputates the leg because he is a sadist and wants to cause the soldier excruciating pain. In this case, the surgeon acts objectively rightly since he saves the soldier's life, but he seems blameworthy for his action since he does it for a vile reason. There are thus cases of praiseworthy objective wrongdoing and cases of blameworthy objective right-doing. So, if the sense of "right" and "wrong" operating in the argument of the previous paragraph is the objective sense, it fails.

However, what if the sense of "right" and "wrong" at work in the argument is the subjective sense? Is the argument any good then? No. I think one can also challenge the link between praiseworthiness/blameworthiness and subjective rightness/wrongness. While there may be a tight connection between subjective wrongdoing and blameworthiness, I think that subjective right-doing can come apart from praiseworthiness. Huck is a good example of this. Huck is surely praiseworthy for his action. But I've argued that given Huck's beliefs and evidence, it's difficult to maintain that his action is subjectively right. We've seen that consideration of plausible theories of subjective rightness supports this conclusion. Thus, my suggestion that a person can be both praiseworthy and blameworthy for an action doesn't necessarily imply that the person's action is both subjectively right and subjectively wrong. And even if it did, it's

³⁰ For similar cases, see Zimmerman (1997).

not clear to me that this would be a devastating result. While I can see why one would want to deny that an action can be both objectively right and objectively wrong, it's less clear to me why one would want to deny that an action can be both subjectively right and subjectively wrong. After all, subjective rightness and wrongness are determined by an agent's mental state, in particular her beliefs or evidence. And why couldn't there be a possible configuration of mental states according to which a person's action is both subjectively right and subjectively wrong? Suppose, for instance, that a utilitarian has inconsistent beliefs and holds (perhaps consciously) that one of her actions is best, but she also holds (perhaps unconsciously) that it is not. If that could happen, and I don't see why it couldn't, then why shouldn't we say that the action in question is, for the person, both subjectively right and subjectively wrong? I am therefore not convinced that the argument above is any better if the sense of "right" and "wrong" at issue is the subjective sense rather than the objective sense.

However, for those who still feel uneasy about the contradictory air of my proposal, here's a way re-characterizing it so that it doesn't have even the appearance of being contradictory. I have assumed throughout this chapter that a person is praise/blameworthy for performing *actions*. And my suggestion is that a person can be both praiseworthy and blameworthy for performing one and the same action. This happens in cases like the Huck case, where a person performs an action from a good motive but believes the action is wrong. In contrast to this general picture, however, one might deny that we are ever praise/blameworthy for performing actions *simpliciter*. One might hold instead that we are praise/blameworthy for performing *complexes* of things that involve actions, and one might take this chapter to have identified two interesting

complexes that we are praise/blameworthy for. In particular, we can be praise/blameworthy for acting from a certain motive [action + motive], and we can be praise/blameworthy for acting in light of a certain belief [action + belief]. Applied to the Huck case, the idea is to (a) deny that Huck is praiseworthy and blameworthy for performing one and the same action and (b) maintain instead that Huck is praiseworthy for one thing (acting from a good motive) and blameworthy for quite a different thing (acting in spite of his belief that his action is wrong). And surely there is nothing contradictory about maintaining that.³¹

Of course, in denying that people can be praise/blameworthy for performing actions, this re-characterization of my view denies an assumption commonly made in the literature. But perhaps this assumption is unwarranted. Indeed, it's difficult to think of good independent reasons for thinking that people are praise/blameworthy for actions rather than for complexes of things that involve actions. Here, for instance, is one potential reason for thinking this. One might claim that it's surely the case that we are obliged/permitted/forbidden to perform actions (rather than complexes of things, such as an action + a motive). However, we should hold that the type of thing that we are obliged/permitted/forbidden to perform is identical to the type of thing (or, at least, one of the types of things) that we are praise/blameworthy for. Thus we are praise/blameworthy for performing actions. The problem with this argument, however, is with its premise that the type of thing that we are obliged/permitted/forbidden to perform is identical to the type of thing that we are praise/blameworthy for. What is the rationale for this premise supposed to be? It's hard to see a good answer to this question.

³¹ Thanks to Fred Feldman for suggesting this way of re-characterizing my position.

After all, the deontic mode of assessment (which has to do with obligation/permissibility/wrongness) is a different mode of assessment from the appraisability mode of assessment (which has to do with praiseworthiness/blameworthiness). Given this, why think that the objects of these two modes of assessment must align?

So, although I see no decisive reason to re-characterize the proposal I have offered in this chapter in the way currently under consideration, I also see no great reason to resist that re-characterization either. It is, of course, possible that further inquiry into this matter will tip the balance one way or the other. However, for my purposes, this is largely inconsequential, for no matter which way the balance is tipped, I think we will have a plausible way of building on Ross's theory of moral goodness from *R&G* so it can adequately handle all of the cases presented in this chapter.

8.7. Conclusion

In this chapter, I have claimed that Ross's theory of moral goodness from *R&G* (namely, RMG/RMB) is superior to both Kant's theory and the Coincident Reasons Thesis. However, Ross's theory is subject to the nepotism problem, and I spent much of the chapter discussing how it should be solved. First, I argued that the problem cannot be solved either by transforming Ross's theory to RMG_{FE}/RMB_{FE} or to RMG₊/RMB₊. Then I turned to my own solution, which takes Ross's theory of moral goodness from *R&G* as a starting point and builds on it. In particular, it acknowledges that a person is praiseworthy (blameworthy) for performing an action when the person does it from a good (bad) motive, but it maintains that there are other things that can make a person

praiseworthy (blameworthy) as well, e.g., performing an action in the belief that it ought (ought not) to be done, and exhibiting a virtue (vice) while performing the action. This view, I claim, yields the intuitively correct conclusions about all of the cases discussed in this chapter, in particular the Huck case, the nepotism case, and the Brenda case. However, I acknowledge that I have merely presented a sketch of the theory here; more will need to be said before it can be considered fully plausible.

8.8. Appendix: Markovits on Fundamental and Non-Instrumental Reasons

Recall the first problem for CRT that I mentioned: the theory is designed to make the intuitively correct conclusions about the Huck and Charlie cases, but at first glance, it doesn't seem to make those conclusions. After all, the motive that Huck and Charlie act upon is surely not the motive to maximize utility, or to treat Humanity as an end in itself, or to do what a virtuous person would do. Huck's and Charlie's motives are surely much simpler than that. But this means that if utilitarianism, Kantianism, or virtue ethics is true, then Huck's and Charlie's actions have no moral worth on CRT. Similar things can be said if any of the other standard theories in normative ethics is true. Above, I suggested a way for a proponent of CRT to solve this problem: appeal to different levels of moral explanation. While this solution is similar to the one Markovits offers, it doesn't appear to be exactly the same. In this appendix, I discuss some of the details of Markovits's solution and raise some problems for it.

To solve the present problem for CRT, Markovits draws a distinction that is supposed to be analogous to the distinction Korsgaard makes between intrinsic and non-instrumental value. According to Markovits, the idea that Korsgaard has in mind is this.

To say that something is intrinsically valuable is to say that it “doesn’t get its value from the value of any other thing but rather, as Korsgaard puts it, ‘has its goodness in itself’” (2010, 228). On the other hand, to “describe something as noninstrumentally valuable—valuable as an end—is to say something about the *way* we value it: we value it for its own sake” (228). So, the distinction Korsgaard has in mind, at least as Markovits characterizes it, seems fairly clear. An intrinsically valuable thing is a thing that is valuable for its own sake; a non-instrumentally valuable thing is a thing that we value for its own sake. Something could thus be non-instrumentally valuable without being intrinsically valuable. Such a thing would be one that we value for its own sake, but it is not in fact good in itself.³² Markovits points out that while utilitarians and Kantians disagree about what things are intrinsically valuable, they can agree, to a large extent, about what is non-instrumentally valuable. For instance, utilitarians think that happiness (or perhaps pleasure) is the only thing that’s intrinsically good, whereas Kantians think that the good will is the only thing that’s intrinsically good; however, both parties can agree that happiness is non-instrumentally good: we value it for its own sake (229).

Markovits (2010, 229) claims that similar things can be said about reasons for acting. On the one hand, there are fundamental (or intrinsic) reasons for acting. This is the type of reason that utilitarians and Kantians disagree about. Utilitarians think that the most fundamental reason for acting has to do with maximizing utility; Kantians think that

³² Could something be intrinsically valuable without being non-instrumentally valuable? It seems so. Such a thing would be good in itself but one that we value only for the sake of other things. However, surprisingly, Markovits claims that this is not possible: “Merely instrumentally valuable things or actions are never intrinsically (unconditionally) valuable” (2010, 228). It’s unclear to me why she thinks this.

the most fundamental reason for acting has to do with treating Humanity as an end in itself, not as a mere means. However, alongside fundamental (intrinsic) reasons for acting, there are non-instrumental reasons for acting. And while utilitarians and Kantians disagree about the fundamental reasons for acting, they will agree, to a large extent, about the non-instrumental reasons for acting. In particular, they will both (likely) agree that easing someone's pain, helping a friend, and increasing someone's well-being are non-instrumental reasons for acting.

This opens the door to a solution to our problem. Markovits thinks that “We should understand the Coincident Reasons Thesis as pronouncing an action morally worthy whenever the noninstrumental reasons for which it is performed coincide with the noninstrumental reasons that morally justify its performance” (2010, 230; italics omitted). But if we understand the thesis in this manner, then the present problem seems to evaporate. For although Huck and Charlie surely do not perform their actions for the *fundamental* reasons that justify their actions, they do (utilitarians and Kantians can agree) perform their actions for the *non-instrumental* reasons that justify their actions. So, if we understand the Coincident Reasons Thesis in the way that Markovits suggests, it implies that Huck's and Charlie's actions have moral worth after all.

One thing to note at this juncture is that the way Markovits seems to understand the Coincident Reasons Thesis here is conceptually quite different from how I initially presented it. As presented above, the thesis is this:

CRT: An action has moral worth iff the agent performs the action for the reasons why it is morally right (i.e., for the reasons that make it morally right).

However, in the passages just cited, Markovits seems to understand the thesis more like this:

CRT*: An action has moral worth iff the agent performs the action for the non-instrumental reasons that justify its performance.³³

But CRT and CRT* make very different claims, even if we stipulate that the type of reason involved in CRT is the non-instrumental type. There's an important difference between the reason *why* an action is right and the reasons that *justify* its performance. A reason why an action is right is a consideration or feature that makes the action right or grounds its rightness. Such a reason is typically called an "explanatory" reason. Explanatory reasons, however, are quite different from what are commonly referred to as "normative" or "justifying" reasons. A normative reason isn't (necessarily) a consideration that explains why an action is right; rather, it is a consideration that, roughly speaking, counts in favor of the action.³⁴ There is thus an important conceptual difference between explanatory reasons and normative reasons. Sometimes when Markovits talks about the Coincident Reasons Thesis, it looks to be a thesis about explanatory reasons; other times, it looks to be a thesis about normative reasons. Markovits slides back and forth between these two ways of understanding the thesis.

In response to this, I suspect that Markovits would say that these two ways of understanding the thesis are equivalent, for a reason is justifying iff it is explanatory. In

³³ Here's another way of stating this: an action has moral worth just in case the agent's motivating reasons coincide with the reasons that justify its performance. Once again, as Markovits makes clear, the motives involved in CRT* are the agent's non-instrumental motives, not her instrumental ones.

³⁴ For more on this distinction, see Schroeder (2007, 10-15).

other words, I suspect she'd say that a consideration counts in favor of an action iff it explains why the action is right. However, it's not clear to me that this claim is true. In fact, Ross seems to deny it. As we've seen, he suggests that the fact that an action fulfills a promise is a consideration that counts in its favor (i.e., is a *prima facie* duty), but when the action is right, it's right not because it keeps a promise, but rather, because it maximizes the balance of total *prima facie* rightness over total *prima facie* wrongness (see sec. 1.4). I, for one, am sympathetic with this type of view.

Markovits might say that Ross is talking here about fundamental reasons, not non-instrumental ones, and while it's not the case that a reason is fundamentally justifying iff it is fundamentally explanatory, it *is* the case that a reason is non-instrumentally justifying iff it is non-instrumentally explanatory. But is this true? It's hard for me to tell, for it's not clear to me what exactly Markovits has in mind when she distinguishes between fundamental and non-instrumental justifying reasons, nor is it clear to me what the difference between fundamental and non-instrumental explanatory reasons is supposed to be. As we've seen, Markovits suggests that the distinction between fundamental and non-instrumental justifying reasons is analogous to Korsgaard's distinction between intrinsic and non-instrumental value, which amounts to the distinction between things that are valuable in themselves and things that we value for themselves. Given this, one might think that the difference between fundamental and non-instrumental justifying reasons is this: fundamental justifying reasons are considerations that count, in themselves, in favor of an action, whereas non-instrumental justifying reasons are considerations that *we count*, in themselves, in favor of an action. Thus, like non-instrumentally valuable things, non-instrumental justifying reasons

crucially depend on the manner in which we regard something, which is not the case for intrinsically valuable things or fundamental justifying reasons. If non-instrumental justifying reasons are understood in this way, then CRT* amounts to the claim that an action has worth iff the agent performs it for the considerations that she counts, in themselves, in favor of the action.

However, I doubt that Markovits intends CRT* to be understood in this way. This interpretation of CRT* has some uncomfortable implications. Hitler may well have taken the fact that an action of his exterminates the Jews to count (in itself) in favor of the action. The present understanding of CRT* would then appear to imply that Hitler's action has moral worth, which I doubt Markovits will want to maintain. Also, this understanding of CRT* seems inconsistent with the spirit of the thesis, which is that actions have moral worth only if they are done for the reasons that, *in fact*, justify the action/make it right, whether or not the agent *considers* them to justify the action/make it right. The Huck Finn case is supposed to motivate this.

But if the distinction between fundamental and non-instrumental justifying reasons should not be understood in the way just suggested, then what does it amount to? The closest Markovits comes to elucidating the distinction is when she says,

Noninstrumental reasons are those reasons to act that are provided by the ends of our actions that are worth pursuing for their own sake (in the case of normative reasons) or that we pursue for their own sake (in the case of motivating reasons). Thus, as I'm using the term, we have non-instrumental reasons to pursue even purely instrumentally valuable actions—they're provided by the end, valued for its own sake, to which we perform those actions. Our fundamental reasons for acting, normative and motivating, are always noninstrumental—our reasons cannot be instrumental all the way down. (2010, 229)

This passage is difficult to make sense of. Markovits suggests that non-instrumental normative (i.e., justifying) reasons are “provided” by the “ends” of our actions that are worth pursuing for their own sake. So, Markovits evidently thinks that actions have certain ends. However, it’s unclear to me what exactly these ends are supposed to be. Is the end of an action its consequences? Is it its perceived consequences? Is it the motive the action is performed on? Or what? It’s also unclear to me what it means for a reason to be “provided” by these ends. Perhaps this means that the reason exists *because* the ends exist, but I’m not sure. Furthermore, recall that Markovits thinks that utilitarians, Kantians, etc. will agree that there are many non-instrumental justifying reasons, and they will largely agree about what those reasons are. But this may seem difficult to reconcile with Markovits’s characterization of non-instrumental reasons in the above passage. After all, wouldn’t a utilitarian say that the only end worth pursuing for its own sake is the maximization of utility, whereas a Kantian would say that the only end worth pursuing for its own sake is the treatment of Humanity as an end in itself? But since these ends are so different, why think that the reasons they provide significantly overlap? It all depends, I suppose, on what “provide” means in this context. But as I’ve said, it’s not clear to me what exactly that is.

All of this is to say that the fundamental/non-instrumental justifying reason distinction that Markovits appeals to is difficult to grasp. As a result, it’s difficult to determine whether CRT* can in fact handle the Huck and Charlie cases. It also seems to me that the second problem for CRT that I mentioned applies to CRT* as well. The second problem, recall, concerns actions done solely because they are right. I asked us to consider a case of a soldier who heroically dives on a grenade, and his sole (non-

instrumental) motive for doing so is to act rightly. Since the fact that an action is right is not a reason that explains why it is right, CRT implies that the soldier's action has no moral worth, which is implausible. A similar problem arises for CRT*. This is because the fact that an action is right is not a reason that justifies its performance. That is, the fact that an action is right is not a consideration that counts in its favor. This claim is widely accepted (see, e.g., Dancy 2004, 16; Stratton-Lake 2000, chap. 1; 2011, 370-71; Broome 2004, 52-55). Furthermore, I presume that it is true regardless of whether fundamental or non-instrumental justifying reasons are at issue. But if that is the case, then CRT*, like CRT, implies that any action done solely because it is right has no moral worth.

It should be noted that Markovits (2010, 218-19) is aware of this potential problem for her view, but she is untroubled by it. She suggests that an action done solely because it is right is an action done at least in part because of the agent's belief that it is right. When this belief is unjustified, the action has no worth; however, when the belief is justified, the action does have worth. But Markovits thinks that this is compatible with CRT*, for actions done from the (justified) belief that they are right are done for the reasons that justify their performance. But is this true? The crucial question to ask is this: is the fact that an agent believes (justifiably) that an action of hers is right a consideration that counts in its favor? It's not at all clear to me that it is. It's much more plausible to hold that the facts in *virtue of which* the agent justifiably believes that her action is right justify its performance. So, for instance, it's more plausible to hold—as Markovits, at times, suggests—that the fact that a person has good evidence that her action is best counts in favor of the action's performance, as does the fact that the person

has been advised to perform the action by a reliable authority. But, crucially, I'm *not* supposing that when our soldier dives on the grenade, his (non-instrumental) motive is to do what's best, or what the moral authorities tell him he should do. Rather, I'm supposing that the soldier's sole (non-instrumental) motive is to do the right thing. And it seems to me rather implausible—or, at least, highly contentious—to hold that this motive is a consideration that counts in favor of the soldier's action, even if we assume that the motive is his (justified) belief that his action is right. Markovits's response to the present objection thus seems inconclusive at best.

Moreover, CRT* has additional problems. First, a minor one: I think the view is slightly misstated. CRT* says that an action has moral worth iff it is performed for *the* (that is, *all of the*) non-instrumental reasons that justify its performance. But there will presumably be lots of non-instrumental reasons that justify an action, and Markovits will surely want to say that an action has worth if it is performed for some, but not all, of those reasons. In fact, Markovits makes this clear in the final section of her paper (2010, 237-41). There she suggests that the degree to which an action has moral worth is a function of the degree to which the agent's motivating reasons coincide with the reasons justifying the action. Thus, an action that is performed for some, but not all, of the non-instrumental reasons that justify it will have partial, but not full, moral worth. So, I think we need to either change “an action has moral worth” on CRT*'s left-hand side to “an action has *full* moral worth,” or change “the agent performs the action for the non-instrumental reasons” on CRT*'s right-hand side to “the agent performs the action for *some of the* non-instrumental reasons.”

Second, CRT* is subject to the nepotism problem. Recall that in the nepotism case, Ned hires his relative out of a good motive (namely, to benefit his relative) even though he thinks his action is wrong since there are better candidates. Plausibly, the fact that an action benefits someone (non-instrumentally) justifies the action; the fact that an action benefits someone presumably counts in favor of the action's performance. But if that's the case, then CRT* appears to imply that Ned's action has moral worth. Now, as I've suggested, I *do* think that we should say that Ned's action has some moral worth (i.e., he's somewhat praiseworthy for his action). However, Ned is also clearly blameworthy for what he's done. In fact, Ned seems much *more* blameworthy for his deed than he seems praiseworthy for it. It's not clear to me how a proponent of CRT* can capture the intuitive moral unworth of Ned's action.

Finally, Gert (2012) raises an interesting problem for CRT*. He argues that motivating reasons have only one dimension of strength, but normative reasons have two: justifying strength and requiring strength. It therefore makes little sense to talk of motivating reasons "coinciding" with normative reasons. As Gert nicely puts it, "As a result, one's motivating reasons can no more coincide with the normative reasons of relevance to one's action than a line can coincide with a square, or a square with a cube" (612).

These problems, it should be noted, are ones that the view I have proposed in this chapter avoids. My view makes the intuitively correct conclusions about the Huck case without having to appeal to any mysterious distinctions, such as the distinction between fundamental and non-instrumental reasons. My view also implies that an action is praiseworthy when it is done solely because it is right (at least if we assume, as Ross

does, that the motive of duty is a good motive). Moreover, I've argued that my view yields the intuitively correct implications about the nepotism case. And since it does not appeal to normative reasons or their strengths, it also avoids Gert's objection. So, it seems to me that Ross's theory, or at least the extension of it that I have proposed, is still preferable to the Coincident Reasons Thesis, even if the thesis is understood as CRT* instead of as CRT.

BIBLIOGRAPHY

- Arpaly, Nomy. 2002. "Moral Worth." *The Journal of Philosophy* 99: 223-45.
- . 2003. *Unprincipled Virtue*. Oxford: Oxford University Press.
- Audi, Robert. 1996. "Intuitionism, Pluralism, and the Foundations of Ethics." In *Moral Knowledge: New Readings in Moral Epistemology*, edited by Walter Sinnott-Armstrong and Mark Timmons, 101-36. Oxford: Oxford University Press.
- . 2004. *The Good in the Right: A Theory of Intuition and Intrinsic Value*. Princeton: Princeton University Press.
- Bales, R. Eugene. 1971. "Act-Utilitarianism: Account of Right-Making Characteristics or Decision-Making Procedure?" *American Philosophical Quarterly* 8: 257-65.
- Bennett, Jonathan. 1974. "The Conscience of Huckleberry Finn." *Philosophy* 49: 123-34.
- Bergström, Lars. 1966. *The Alternatives and Consequences of Actions*. Stockholm, Sweden: Alqvist & Wiksell.
- Berker, Selim. 2007. "Particular Reasons." *Ethics* 118: 109-39.
- Brandt, Richard B. 1959. *Ethical Theory: The Problems of Normative and Critical Ethics*. Englewood Cliffs, NJ: Prentice-Hall.
- Broome, John. 2004. "Reasons." In *Reason and Value: Themes from the Moral Philosophy of Joseph Raz*, edited by R. J. Wallace, Philip Petit, Samuel Scheffler, and Michael Smith, 28-55. Oxford: Clarendon Press.
- Castañeda, Hector-Neri. 1968. "A Problem for Utilitarianism." *Analysis* 28: 141-42.
- Cullity, Garrett. 2002. "Particularism and Presumptive Reasons." *Proceedings of the Aristotelian Society* 76: 169-90.
- Dancy, Jonathan. 1983. "Ethical Particularism and Morally Relevant Properties." *Mind* 92: 530-47.
- . 1991a. "An Ethic of Prima Facie Duties." In *A Companion to Ethics*, edited by Peter Singer, 219-29. Malden, MA: Blackwell.

- . 1991b. "Intuitionism." In *A Companion to Ethics*, edited by Peter Singer, 411-20. Malden, MA: Blackwell.
- . 1993. *Moral Reasons*. Oxford: Blackwell.
- . 2000. *Practical Reality*. Oxford: Oxford University Press.
- . 2004a. "Enticing Reasons." In *Reason and Value: Themes from the Moral Philosophy of Joseph Raz*, edited by R. J. Wallace, Philip Petit, Samuel Scheffler, and Michael Smith, 91-118. Oxford: Clarendon Press.
- . 2004b. *Ethics without Principles*. Oxford: Clarendon Press.
- Davidson, Donald. 1973. "Freedom to Act." In *Essays on Freedom of Action*, edited by Ted Honderich, 139-56. London: Routledge & Kegan Paul.
- Dean, Richard. 2009. "The Formula of Humanity as an End in Itself." In *The Blackwell Guide to Kant's Ethics*, edited by Thomas E. Hill Jr., 83-101. Malden, MA: Wiley-Blackwell.
- DePaul, Michael R. 1988. "The Problem of the Criterion and Coherence Methods in Ethics." *Canadian Journal of Philosophy* 18: 67-86.
- Dorsey, Dale. 2012. "Objective Morality, Subjective Morality and the Explanatory Question." *Journal of Ethics and Social Philosophy* 6 (3): 1-24.
- Elga, Adam. 2010. "Subjective Probabilities Should Be Sharp." *Philosophers' Imprint* 10 (5): 1-11.
- Feldman, Fred. 1975. "World Utilitarianism." In *Analysis and Metaphysics: Essays in Honor of R. M. Chisholm*, edited by Keith Lehrer, 255-71. Dordrecht, Holland: D. Reidel.
- . 1978. *Introductory Ethics*. Englewood Cliffs, NJ: Prentice-Hall.
- . 1986. *Doing the Best We Can: An Essay in Informal Deontic Logic*. Dordrecht, Holland: D. Reidel.
- . 2006. "Actual Utility, the Objection from Impracticality, and the Move to Expected Utility." *Philosophical Studies* 129: 49-79.

- . 2012. “True and Useful: On the Structure of a Two Level Normative Theory.” *Utilitas* 24: 151-71.
- . In Progress. *A Desertist Theory of Justice*.
- Frazier, Robert L. 1994. “Act Utilitarianism and Decision Procedures.” *Utilitas* 6: 43-53.
- Gaut, Berys. 2002. “Justifying Moral Pluralism.” In *Ethical Intuitionism: Re-evaluations*, edited by Philip Stratton-Lake, 137-60. Oxford: Clarendon Press.
- Gert, Joshua. 2003. “Requiring and Justifying: Two Dimensions of Normative Strength.” *Erkenntnis* 59: 5-36.
- . 2004. *Brute Rationality: Normativity and Human Action*. Cambridge: Cambridge University Press.
- . 2012. “Moral Worth, Supererogation, and the Justifying/Requiring Distinction.” *The Philosophical Review* 121: 611-18.
- Gibbard, Allan. 2003. “Normative Properties.” *The Southern Journal of Philosophy* 41: 141-57.
- . 2005. “Truth and Correct Belief.” *Philosophical Issues* 15: 338-50.
- Graham, Peter A. 2010. “In Defense of Objectivism about Moral Obligation.” *Ethics* 121: 88-115.
- Hájek, Alan. 2012. “Interpretations of Probability.” In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta.
<http://plato.stanford.edu/archives/win2012/entries/probability-interpret/>.
- Haji, Ishtiyaque. 1997. “An Epistemic Dimension of Blameworthiness.” *Philosophy and Phenomenological Research* 57: 523-44.
- . 1998. *Moral Appraisability: Puzzles, Proposals, and Perplexities*. Oxford: Oxford University Press.
- Hedden, Brian. 2012. “Options and the Subjective Ought.” *Philosophical Studies* 158: 343-60.

- Henson, Richard G. 1979. "What Kant Might Have Said: Moral Worth and the Overdetermination of Dutiful Action." *The Philosophical Review* 88: 39-54.
- Herman, Barbara. 1981. "On the Value of Acting from the Motive of Duty." *The Philosophical Review* 90: 359-82.
- Hill, Thomas E., Jr. 1980. "Humanity as an End in Itself." *Ethics* 91: 84-99.
- Hooker, Brad. 1996. "Ross-Style Pluralism Versus Rule-Consequentialism." *Mind* 105: 531-52.
- . 2000. *Ideal Code, Real World: A Rule-Consequentialist Theory of Morality*. Oxford: Oxford University Press.
- . 2008. "Moral Particularism and the Real World." In *Challenging Moral Particularism*, edited by Mark Lance, Matjaz Potrc, and Vojko Strahovnik, 12-30. New York: Routledge.
- Howard-Snyder, Frances. 2005. "It's the Thought that Counts." *Utilitas* 17: 265-81.
- Hurka, Thomas. 2006. "Virtuous Act, Virtuous Dispositions." *Analysis* 66: 69-76.
- . 2011. "Common Themes from Sidgwick to Ewing." In *Underivative Duty: British Moral Philosophers from Sidgwick to Ewing*, edited by Thomas Hurka, 6-25. Oxford: Clarendon Press.
- Hursthouse, Rosalind. 1999. *On Virtue Ethics*. Oxford: Oxford University Press.
- Jackson, Frank. 1991. "Decision-Theoretic Consequentialism and the Nearest and Dearest Objection." *Ethics* 101: 461-82.
- Jackson, Frank, Philip Pettit, and Michael Smith. 2000. "Ethical Particularism and Patterns." In *Moral Particularism*, edited by Brad Hooker and Margaret Olivia Little, 79-99. Oxford: Clarendon Press.
- Joyce, James M. 2010. "A Defense of Imprecise Credences in Inference and Decision Making." *Philosophical Perspectives* 24: 281-323.
- Kant, Immanuel. (1785) 1997. *Groundwork of the Metaphysics of Morals*. Edited by Mary Gregor. Reprint, Cambridge: Cambridge University Press.
- Kagan, Shelly. 1989. *The Limits of Morality*. Oxford: Clarendon Press.

- Kleist, Chad. 2009. "Huck Finn the Inverse Akratic: Empathy and Justice." *Ethical Theory and Moral Practice* 12: 257-66.
- Korsgaard, Christine M. 1996. *Creating the Kingdom of Ends*. Cambridge: Cambridge University Press.
- Lance, Mark Norris, and Margaret Olivia Little. 2006a. "Defending Moral Particularism." In *Contemporary Debates in Moral Theory*, edited by James Dreier, 305-21. Malden, MA: Blackwell.
- . 2006b. "Particularism and Antittheory." In *The Oxford Handbook of Ethical Theory*, edited by David Copp, 567-94. Oxford: Oxford University Press.
- . 2008. "From Particularism to Defeasibility in Ethics." In *Challenging Moral Particularism*, edited by Mark Lance, Matjaz Potrc, and Vojko Strahovnik, 53-74. New York: Routledge.
- Leibowitz, Uri D. 2009. "A Defense of a Particularist Research Program." *Ethical Theory and Moral Practice* 12: 181-99.
- Lenman, James. 2000. "Consequentialism and Cluelessness." *Philosophy and Public Affairs* 29: 342-70.
- Little, Margaret Olivia. 2000. "Moral Generalities Revisited." In *Moral Particularism*, edited by Brad Hooker and Margaret Olivia Little, 276-304. Oxford: Clarendon Press.
- Markosian, Ned. 2009. "Rossian Minimalism." *Journal of Ethics and Social Philosophy* 4 (1): 1-17.
- Markovits, Julia. 2010. "Acting for the Right Reasons." *The Philosophical Review* 119: 201-42.
- Mason, Elinor. 2013. "Objectivism and Prospectivism about Rightness." *Journal of Ethics and Social Philosophy* 7 (2): 1-21.
- Matthews, Gareth. 1972. "Senses and Kinds." *The Journal of Philosophy* 69: 149-57.
- McCann, Hugh. 2007. "Metaethical Reflections on Robert Audi's Moral Intuitionism." In *Rationality and the Good*, edited Mark Timmons, John Greco, and Alfred R. Mele, 40-53. Oxford: Oxford University Press.

- McConnell, Terrance. 1988. "Ross on Duty and Ignorance." *History of Philosophy Quarterly* 5: 79-95.
- McKeever, Sean, and Michael Ridge. 2005. "The Many Moral Particularisms." *Canadian Journal of Philosophy* 35: 83-106.
- . 2006. *Principled Ethics: Generalism as a Regulative Ideal*. Oxford: Clarendon Press.
- . 2008. "Preempting Principles: Recent Debates in Moral Particularism." *Philosophy Compass* 3: 1177-92.
- McLeod, Owen. 2001. "Just Plain "Ought." *The Journal of Ethics* 5: 269-91.
- . 2013. "Desert." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. <http://plato.stanford.edu/archives/win2013/entries/desert/>.
- McNaughton, David. 1988. *Moral Vision*. Malden, MA: Blackwell.
- . 1996. "An Unconnected Heap of Duties." *The Philosophical Quarterly* 46: 433-47.
- . 2000. "Intuitionism." In *The Blackwell Guide to Ethical Theory*, edited by Hugh LaFollette, 268-87. Malden, MA: Blackwell.
- Mellor, D. H. 2005. *Probability: A Philosophical Introduction*. New York: Routledge.
- Miller, Dale E. 2003. "Actual-Consequence Act Utilitarianism and the Best Possible Humans." *Ratio* 16: 49-62.
- Montague, Phillip. 2000. "Why Rule Consequentialism is not Superior to Ross-Style Pluralism." In *Morality, Rules, and Consequences: A Critical Reader*, edited by Brad Hooker, Elinor Mason, and Dale E. Miller, 203-11. New York: Rowman & Littlefield.
- Moore, G. E. (1903) 1993. *Principia Ethica*. Rev. ed. Edited by Thomas Baldwin. Cambridge: Cambridge University Press.
- . (1912) 2005. *Ethics*. Edited by W. H. Shaw. Reprint, Oxford: Clarendon Press.

- Olsaretti, Serena. 2003. "Introduction: Debating Desert and Justice." In *Desert and Justice*, edited by Serena Olsaretti, 1-24. Oxford: Clarendon Press.
- Parfit, Derek. 1984. *Reasons and Persons*. Oxford: Clarendon Press.
- . 2011. *On What Matters*. Vol. 1. Edited and introduced by Samuel Scheffler. Oxford: Oxford University Press.
- Pogge, Thomas W. 1998. "The Categorical Imperative." In *Kant's "Groundwork of the Metaphysics of Morals": Critical Essays*, edited by Paul Guyer, 189-213. Lanham, MD: Rowman & Littlefield.
- Portmore, Douglas W. 2008. "Dual-Ranking Act-Consequentialism." *Philosophical Studies* 138: 409-27.
- . 2011. *Commonsense Consequentialism: Wherein Morality Meets Rationality*. Oxford: Oxford University Press.
- Prichard, H. A. (1932) 2002. "Duty and Ignorance of Fact." In *Moral Writings*, edited by Jim MacAdam, 85-110. Oxford: Oxford University Press.
- Regan, Donald. 1980. *Utilitarianism and Co-Operation*. Oxford: Clarendon Press.
- Robinson, Michael. 2010. "Are Some Prima Facie Duties More Binding than Others?" *Utilitas* 22: 26-32.
- Rawls, John. 1999. *A Theory of Justice*. Rev. ed. Cambridge, MA: Belknap Press.
- Ross, W. D. 1928-29. "The Nature of Morally Good Action." *Proceedings of the Aristotelian Society* 29: 251-74.
- . 1930. *The Right and the Good*. Oxford: Clarendon Press.
- . 1939. *The Foundations of Ethics*. Oxford: Clarendon Press.
- Russell, Bertrand. 1966. "The Elements of Ethics." In *Philosophical Essays*, 13-59. New York: Simon and Schuster.
- Scanlon, T. M. 1998. *What We Owe to Each Other*. Cambridge, MA: Harvard University Press.
- Schroeder, Mark. 2007. *Slaves of the Passions*. Oxford: Oxford University Press.

- . 2008. "Having Reasons." *Philosophical Studies* 139: 57-71.
- Shafer-Landau, Russ. 1997. "Moral Rules." *Ethics* 107: 584-611.
- . 2012. *The Fundamentals of Ethics*. 2nd ed. Oxford: Oxford University Press.
- Sidgwick, Henry. 1907. *The Methods of Ethics*. 7th ed. New York: Macmillan.
- Singer, Peter. 1972. "Famine, Affluence, and Morality." *Philosophy & Public Affairs* 1: 229-43.
- Slote, Michael. 1984. "Morality and Self-Other Asymmetry." *The Journal of Philosophy* 81: 179-92.
- Smith, Holly M. 1976. "Dated Rightness and Moral Imperfection." *The Philosophical Review* 85: 449-87.
- . 1978. "Doing the Best One Can." In *Values and Morals: Essays in Honor of William Frankena, Charles Stevenson, and Richard Brandt*, edited by Alvin I. Goldman and Jaegwon Kim, 185-214. Dordrecht, Holland: D. Reidel.
- . 1988. "Making Moral Decisions." *Nous* 22: 89-108.
- . 1997. "A Paradox of Promising." *The Philosophical Review* 106: 153-96.
- . 2010. "Subjective Rightness." *Social Philosophy and Policy* 27: 64-110.
- . 2012. "Using Moral Principles to Guide Decisions." *Philosophical Issues* 22: 369-86.
- Smith, Michael. 1994. *The Moral Problem*. Malden, MA: Blackwell.
- Stocker, Michael. 1976. "Agent and Other: Against Ethical Universalism." *Australasian Journal of philosophy* 54: 206-20.
- Stratton-Lake, Philip. 2000. *Kant, Duty and Moral Worth*. London: Routledge.
- . 2002. Introduction to *The Right and the Good*, by W. D. Ross, edited by Philip Stratton-Lake, ix-lviii. Oxford: Clarendon Press.
- . 2011a. "Eliminativism about Derivative Prima Facie Duties." In *Underivative Duty: British Moral Philosophers from Sidgwick to Ewing*, edited by Thomas Hurka, 146-65. Oxford: Clarendon Press.

- Strawson, P. F. 1962. "Freedom and Resentment." *Proceedings of the British Academy* 47: 1-25.
- . 2011b. "Recalcitrant Pluralism." *Ratio* 24: 364-383.
- Taylor, Craig. "Huck Finn, Moral Reasons and Sympathy." *Philosophy* 87: 583-93.
- Thomson, Judith Jarvis. 1986. "Imposing Risks." In *Rights, Restitution, and Risks*, edited by William Parent, 173-91. Cambridge, MA: Harvard University Press.
- Timmons, Mark. 2002. *Moral Theory: An Introduction*. Lanham, MD: Rowman & Littlefield.
- . 2006. "The Categorical Imperative and Universalizability." In *Groundwork for the Metaphysics of Morals*, edited by Christoph Horn and Dieter Shönecker, 158-99. Berlin: Walter de Gruyter.
- Twain, Mark. 1885. *Adventures of Huckleberry Finn*. New York: Charles L. Webster.
- Väyrynen, Pekka. 2006. "Moral Generalism: Enjoy in Moderation." *Ethics* 116: 707-41.
- . 2009. "A Theory of Hedged Moral Principles." In vol. 4 of *Oxford Studies in Metaethics*, edited by Russ Shafer-Landau, 91-132. Oxford: Oxford University Press.
- Wallace, R. Jay. 1994. *Moral Responsibility and the Moral Sentiments*. Cambridge, MA: Harvard University Press.
- White, Roger. 2010. "Evidential Symmetry and Mushy Credence." In *Oxford Studies in Epistemology*, edited by Tamar Szabó Gendler and John Hawthorne, 161-87. Oxford: Oxford University Press.
- Wood, Allen. 1998. "Humanity as an End in Itself." In *Kant's "Groundwork of the Metaphysics of Morals": Critical Essays*, edited by Paul Guyer, 165-87. Lanham, MD: Rowman & Littlefield.
- Zimmerman, Michael J. 1996. *The Concept of Moral Obligation*. Cambridge: Cambridge University Press.
- . 1997. "A Plea for Accuses." *American Philosophical Quarterly* 34: 229-43.

- . 2004. "Another Plea for Excuses." *American Philosophical Quarterly* 41: 259-66.
- . 2006. "Is Moral Obligation Objective or Subjective?" *Utilitas* 18: 329-61.
- . 2008. *Living with Uncertainty: The Moral Significance of Ignorance*. Cambridge: Cambridge University Press.