

RATIONAL CAPACITIES AND THE PRACTICE OF BLAME
A SKEPTICAL ARGUMENT

A Thesis

by

ZACHARY CHARLES BACHMAN

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of
MASTER OF ARTS

May 2011

Major Subject: Philosophy

Rational Capacities and the Practice of Blame: A Skeptical Argument

Copyright 2011 Zachary Charles Bachman

RATIONAL CAPACITIES AND THE PRACTICE OF BLAME
A SKEPTICAL ARGUMENT

A Thesis

by

ZACHARY CHARLES BACHMAN

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF ARTS

Approved by:

Chair of Committee,	Hugh McCann
Committee Members,	Linda Radzik
	Brandon Schmeichel
Head of Department,	Daniel Conway

May 2011

Major Subject: Philosophy

ABSTRACT

Rational Capacities and the Practice of Blame:

A Skeptical Argument. (May 2011)

Zachary Charles Bachman, B.A., University of California, San Diego

Chair of Advisory Committee: Dr. Hugh McCann

This thesis investigates the relationship between our rational capacities and the norms that govern our practice of blame. The conclusion it reaches is rather shocking: *it is impossible to satisfy the conditions of blameworthiness*. The argument that reaches this conclusion is what I call an *internal* criticism. Unlike other skeptical arguments about moral responsibility, the one advanced in this thesis does not depend on any metaphysical theses external to the theory of blame.

The thesis begins by looking at a position I call *rational capacity compatibilism* (RCC). My interest in RCC concerns the fact that it has done more than any other theory of responsibility to set out the relationship between our rational capacities and the practice of blame. I use the most well developed RCC view—that of R. Jay Wallace—as the backdrop for the skeptical argument that follows

I next defend a recent argument advanced by Gideon Rosen according to which an agent cannot be blameworthy for a given act if *akrasia* does not occur somewhere in the act's etiology. This serves as the first major premise in my skeptical argument.

Next, I turn to the second major premise of my argument, which is comprised of two controversial claims. The first is that *akrasia* results from a failure in one's rational capacities. The second is that an agent cannot be blameworthy for committing any act that results from a failure in his or her rational capacities. Together, these two claims produce the following premise: when an agent acts *akratically* she cannot be blameworthy for that act.

Now, for any given act, either *akrasia* occurs in that act's etiology or it does not. If it does not, then the agent in question is not blameworthy (first premise); but if *akrasia* *does* occur in the act's etiology, then the agent in question is still not blameworthy (second premise). It follows that for any given act, the agent who performs that act cannot be blameworthy for so acting.

I end with a brief discussion of what I call "the moral up-shot" of my skeptical argument: what does a world without blame look like? I suggest, contra the main party line (often associated with P.F. Strawson), that blame is not a requirement for significant and meaningful interpersonal relationships, nor is it a necessary component of morality.

DEDICATION

To Richard S., Chuck W., Plumber Doug, Saint Sam, Ray, Scott B., Nate W., and
all the members of the Ocean Beach Clubhouse.

Thank you for saving my life.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my parents: Mary Livingstone, Gary Bachman, Matt Livingstone, and Joanne Basta. Not once have they ever tried to discourage me from pursuing my desire to be an academic philosopher. This is hardly the norm, and I thank them for that.

Thanks are also due to my committee members: Dr. Hugh McCann, Dr. Linda Radzik, and Dr. Brandon Schmeichel. They never once complained about the length and tediousness of the thesis that I handed to them. If anything good is to be found in the pages that follow it is surely due to their helpful comments on previous drafts.

I would also like to thank the graduate students in the Texas A&M philosophy department, past and present. Special thanks are due to Meg Schmitt, Kyle Driggers, Ross Colebrook, Judy Bohr, Tim Ayslworth, David Wright, Robbie Howton, and Anton Kabeshkin. The hours I have spent discussing philosophy with them have been among the most enjoyable hours of my life. I will miss them all dearly.

I would also like to thank Chris Dohna, Kelley Schiffman, Alicia Saake, Kim Brewer, Mason Westfall, Giovanni Macias, and the other regulars at the UCSD philosophy club. It was with them that I truly fell in love with philosophy.

Finally, I would like to again thank Meg Schmitt. I am certain that I would not have survived my time here in College Station if it were not for her. It was with Meg that I bounced off many of my half-baked ideas, and through her warm encouragement that they grew into papers and the thesis you have before you.

TABLE OF CONTENTS

	Page
ABSTRACT	iii
DEDICATION.....	v
ACKNOWLEDGEMENTS	vi
TABLE OF CONTENTS	vii
1. INTRODUCTION: THE PRACTICE OF BLAME	1
2. RATIONAL CAPACITY COMPATIBILISM	6
2.1 A (Very) Brief History of 20 th Century Compatibilism.....	7
2.2 Susan Wolf and “The Reason View”	9
2.3 R. Jay Wallace and Strawsonian Compatibilism.....	13
3. THE SKEPTICAL ARGUMENT I: ROSEN’S SOFT SKEPTICISM	39
3.1 <i>Akrasia</i> : What It Is	39
3.2 Rosen’s Argument	40
3.3 FitzPatrick’s Response.....	51
3.4 Defending Rosen’s Lemma.....	66
4. THE SKEPTICAL ARGUMENT II: THE HARD SKEPTICAL UPSHOT	79
4.1 <i>Akrasia</i> and Irrationality	80
4.2 A Failure in One’s Rational Capacities	92
4.3 The Hard Skeptical Upshot.....	102
5. CONCLUSION: THE MORAL UPSHOT	104
5.1 Summary	104
5.2 Reflections	105
REFERENCES	108
VITA	111

1. INTRODUCTION: THE PRACTICE OF BLAME

It is widely agreed that punishment stands in need of justification. This is because punishment appears to require coercion, and coercion—*prima facie*—is morally wrong. If it is not (always) wrong to punish then a story needs to be told about either why punishment is not actually coercion, or why coercion is not always morally wrong.

I believe that blame is on similar footing. We can see this, I believe, in the immense philosophical literature on moral responsibility. Much of that literature is concerned with the *compatibility* of responsibility with determinism. Ultimately, this is a concern about justification. If determinism proves to be true, our actions are (in a certain sense) unavoidable. But if I cannot do other than what I in fact do, is it really fair to *blame* me for so doing? In deterministic conditions—so the intuition behind the question goes—blaming another for an act is unjustified. Intuitions, of course, vary on this question—one that I am ultimately not concerned with in this project—but that there is a large contingent of individuals who share this intuition is a data point in favor of the claim that blame is in need of justification.

Another data point is the “no blameworthiness without fault principle.” According to this widely held moral principle, an agent is not blameworthy if she has not done anything wrong. If blame did not stand in need of justification there would be no need to restrict the practice to moral wrongs. In other words, there would be no concern

This thesis follows the style of *The Chicago Manual of Style*.

about fairness.

Assuming I am right, and blame, like punishment, *is* in need of justification, there remains a lingering question here, for blame, unlike punishment, does not appear to fall under the category of coercion. So why does it stand in need of justification? The answer to this question is that blame, like coercion (and thus punishment) is *prima facie* harmful. The reason that coercion, and thus punishment, stands in need of justification, is that coercion *harms*. The same is true, I believe, of blame. When we blame people for actions of which they are not in fact blameworthy—as when we blame a person for starting a fire that was really caused by an electrical short—we harm them.

According to the argument that follows, no one is or ever can be blameworthy. Similar conclusions have been reached by others, though their route to this conclusion is far different than mine. A notable example is the argument Derk Pereboom provides in his monograph *Living without Free Will*, which can be characterized as follows:

1. Either determinism is true, or it is not.
2. If determinism is true, then no one has free will.
3. If determinism is false, then we have free will only if agent causation is true.
4. Agent causation, though coherent, is not true (given our best scientific theories).
5. So, we do not have free will. (1,2,3,4)
6. But free will is necessary for blameworthiness.
7. So, we are not blameworthy for our actions (5,6).

The premises of Pereboom's argument are, of course, controversial metaphysical theses, the defense of which, he provides by way of elaborate argument.

One thing to note about Pereboom's conclusion is that it is weaker than the one I draw. According to Pereboom, blameworthiness *is possible*—namely in those worlds where agent causation is true. I, by contrast, argue that blameworthiness is *not possible*. How do I reach this conclusion? (Very) roughly, as follows:

1. For any given agent S and any given action ϕ , if S is blameworthy for ϕ , then either *akrasia* occurs in ϕ 's etiology (either ϕ itself was an *akratic* act or the causal upshot of some *akratic* act) or it does not. (Trivial truth)
2. If *akrasia* does not occur in ϕ 's etiology, then S is not blameworthy for ϕ .
3. If *akrasia* does occur in ϕ 's etiology, then S is not blameworthy for ϕ .
4. So, for any given agent S and any given action ϕ , S is not blameworthy for ϕ .
(1,2,3).

I admit that this argument looks overly simplistic. Pretheoretically, it will likely not be obvious whether premise 2 is true, and premise 3 will appear to be obviously false.

The bulk of this essay is devoted to defending these two premises (I take it that premise 1 does not need defending). Premise 2, which I defend in section 3, is actually a lemma of an argument Gideon Rosen advances in "Skepticism about Moral Responsibility." As I detail in section 3, Rosen does not take this lemma to the conclusion that I do, though he does hint at it: "If there were an independent argument for the thesis that akratic action is never culpable, then we would indeed have a general

argument for the impossibility of responsibility.” However, according to Rosen, “there is no such argument.”¹

In section 4, I provide the argument that Rosen thought non-existent. The argument is fairly complicated and is better left put off until that time. It rests on the relationship between our rational capacities and blameworthiness, a relationship that has been articulated by several prominent philosophers. Chief among them are: Susan Wolf, Peter Strawson, R. Jay Wallace, and John Martin Fischer and Mark Ravizza. According to this position, which I call Rational Capacity Compatibilism (RCC), this relationship is internal to the practice of blame itself. One way to think about this is that, in blaming another, we judge them to be appropriate targets of blame, and what makes a target appropriate is that it possess certain rational capacities. I provide a rough sketch of the RCC position in section 2, with special attention paid to the Strawsonian variety of RCC. Section 4, then, is best understood as read against the back drop of section 2, the main line of argument being put like this: *akratic* action results from a failure in one’s rational capacities, the very failure of which undercuts one’s status of blameworthiness.

What is novel about the argument I provide in this essay is that it is wholly *internal* to the practice of blame itself. That is, unlike with Pereboom’s argument, I do not rely on controversial metaphysical premises concerning the nature of free will and determinism. Rather, my argument depends only on the nature of exempting conditions, which, as we will see in section 1, are part of the notion of blameworthiness itself (as

¹ Gideon Rosen, “Skepticism about Moral Responsibility,” *Philosophical Perspectives* 18 (2004): 295-313.p.308.

free will and determinism are not). I conclude, in section 5, with a brief discussion of what I take the importance of this internal criticism to be.

It will be noted that my skeptical argument relies on the theory of exemptions developed by R. Jay Wallace in *Responsibility and the Moral Sentiments*. As such, my argument might be seen merely as a *reductio* of Wallace's theory and not as an argument for the impossibility of blameworthiness *per se*. While I don't think this is accurate—I think that Wallace's theory accurately captures the practice of blame—space constraints will not allow me to defend myself here. Even so, if my argument *does* only amount to a *reductio* of Wallace's view, I consider this a small win, for it would shift the burden of providing a more adequate theory of blameworthiness back to my opponent.

Before getting the argument underway, I should say a word about my terminology. It has become ubiquitous in the philosophical literature to use the term “blameworthy” synonymously with the phrase “morally responsible.” When I began writing, I was determined not to follow suit. However, when engaging with the literature I found it near impossible to do so while refraining from using the terms synonymously in the way that others do. As such, in what follows, I follow tradition and use the terms synonymously.

2. RATIONAL CAPACITY COMPATIBILISM

In this section, I set out the position I call Rational Capacity Compatibilism (RCC). This position can be characterized by two thesis: (i) when agents are blameworthy it is in virtue of the fact that they possess certain rational capacities; and (ii) the possession of these capacities is compatible with determinism. Philosophers who hold such a position include Susan Wolf, R. Jay Wallace, and John Martin Fischer and Mark Ravizza.

In the sections that follow I ultimately reject the tenability of blameworthiness, and thus reject RCC, though not on incompatibilist grounds (i.e. I don't reject it on the grounds that responsibility is not compatible with determinism). My reason for setting out the RCC position here is that it has done the most in establishing the relationship between the rational and blameworthiness. The argument I advance in section 4 utilizes this relationship to show that the practice of blame is internally unstable.

This section begins with a (very) brief (and rough) history of 20th century compatibilism. The purpose of providing this historical sketch is to put RCC in its dialectical context. I follow that with a brief exegesis of Susan Wolf's Reasons View. The main focus of this section, however, is a position often called *Strawsonian compatibilism*, which is a subcategory of RCC. This position, which develops a theory of moral responsibility out of our reactive attitudes, was first sketched by Peter Strawson in his essay "Freedom and Resentment." But it is in R. Jay Wallace's monograph *Responsibility and the Moral Sentiments* that we find the most sophisticated

development of the RCC position. And it is Wallace's understanding of blameworthiness that I will be using as background against which I mount my skeptical attack in sections 3 and 4.

2.1 A (Very) Brief History of 20th Century Compatibilism

“Compatibilism” is the term that names a large family of solutions to the so-called *problem of determinism*. We can set out this problem accordingly:

THE PROBLEM OF DETERMINISM: If determinism is true, then any given action is caused by an event immediately preceding it. Since this would be true of every event, if determinism is true, then our actions are fully determined by events prior to our birth. But if this is the case, then when I eat a bowl of Post Toastees at t_1 it is not possible that I not eat said bowl of cereal at that time. So if determinism is true, then it is not possible that I not do any given thing that I do. In other words, determinism rules out *the ability to do otherwise*. But if I could not have done otherwise than eat a bowl of Post Toastees at t_1 , then I can't be held responsible for so doing, for I can't be held responsible for doing something that I could not have avoided.² Thus, if determinism is true, then we cannot be held morally responsible for any of our actions.

Traditionally, there have been two solutions proposed to THE PROBLEM OF DETERMINISM. The first is to deny the truth of determinism. This is the so-called *libertarian* solution. I will not be concerned with libertarianism in this essay. The second solution is to deny

² This principle behind this notion is sometimes called the Principle of Alternate Possibilities (PAP): S can be morally responsible for ϕ at t only if S could have done other than ϕ at t .

the implication that determinism rules out morally responsible action. This is the avenue that leads to *compatibilism*. There have been two influential methods of achieving this. The first is to deny that determinism rules out the ability to do otherwise. This strategy has been dubbed “classical compatibilism.” According to the classical compatibilist, the ability to do otherwise ought to receive a *conditional analysis*. That is, to say that *Joan could have done other than ϕ* is just to say that, *if Joan had chosen otherwise, she could have done other than ϕ* . And this, the classical compatibilists maintains, is compatible with the thesis of determinism.

Many have found classical compatibilism to be wanting. The conditional analysis offered is often thought to be vacuous. What good is it to say that I could have done otherwise if I had chosen to, if it were not possible for me (given the truth of determinism) to have chosen otherwise? For if determinism is true, surely I *couldn't* have chosen otherwise.³

The second major compatibilist strategy goes by the moniker “Frankfurt-style compatibilism,” named after Harry Frankfurt, the philosopher who championed the position in several important papers originally published in the late 1960s and early 1970s.⁴ Rather than argue that determinism does not rule out the ability to do otherwise, as the classic compatibilist does, the Frankfurt-style compatibilist argues that alternative possibilities are simply not required for morally responsible action (i.e. PAP is false).

³ Under the Lewis-Stalnaker approach to counterfactual conditional analysis a counterfactual is vacuously true when the antecedent is impossible. Cf. Robert Stalnaker, “A Theory of Conditionals,” in *Studies in Logical Theory*, edited by Nicholas Rescher, American Philosophical Quarterly Monograph Series (Oxford: Basil, 1968).

⁴ Harry Frankfurt, “Alternate Possibilities and Moral Responsibility” in *The Importance of What We Care About*, 1-10 (New York: Cambridge University Press, 1988) and “Freedom of the Will and the Concept of a Person” in *ibid*, 11-25.

We can act, and be responsible for our action, despite the fact that we couldn't do otherwise. What is required for morally responsible action is that our actions be "in line" with our second-order volitions, where this just means that we *affirm* (or would affirm) our action upon reflection; our actions align with who we are and who we want to be. Because of this requirement, Frankfurt-style compatibilism is sometimes called "The Real-Self View."

20th Century compatibilism, then, can be thought of having two branches. On the first branch, we have classical compatibilism, according to which determinism does not threaten morally responsible action because determinism does not rule out the ability to do otherwise. On the second branch we have Frankfurt-style compatibilism, according to which determinism does not threaten morally responsible action because, while determinism *may in fact* rule out the ability to do otherwise, morally responsible action does not require the ability to do otherwise and so determinism is not a threat.

2.2 Susan Wolf and "The Reason View"

By the time Susan Wolf published her monograph, *Freedom within Reason*, two positions stood out as dominating the discussion of solutions to THE PROBLEM OF DETERMINISM. The first is the libertarian view, which requires the denial of determinism—an agent can be considered morally responsible for an act only if she could have done otherwise—along with a rejection of the conditional analysis of the phrase 'could have done otherwise.' While libertarian views vary—for example, some views are agent causal while other views are event causal—the uniting element of all

libertarian views is the commitment to the incompatibility of moral responsibility with determinism. Wolf calls the libertarian position *the autonomy view*.

The second major solution is the Real Self View championed by Frankfurt (and others).⁵ This solution, as mentioned above, rejects that alternative possibilities are necessary for moral responsibility; second-order identification (counterfactual or actual) is sufficient.

Wolf's view can be thought of as meeting these two views halfway. What responsible action requires is the ability to act according to reason.⁶ For Wolf, acting according to reason is something of a dual power.⁷ One must be able to *recognize* or *have access to* moral requirements as well as be able to *act in accordance* with those moral requirements. Thus, in order for me to be morally responsible for setting fire to my neighbor's house, two things must be the case. First, I must be able to recognize or have access to the fact that setting another's house on fire is wrong.⁸ Second, I must be able to *be moved* by this fact into *not* setting the house on fire. This seems intuitively correct; both of these powers are captured in the common excusing conditions of ignorance and insanity.⁹ If I don't have access to the fact that setting another's house on

⁵ For another Real Self View see Gary Watson, "Free Agency," in *Free Will*, Second Edition, ed. Gary Watson (New York: Oxford University Press, 2003), 337-351. The main difference between Frankfurt and Watson's brand of Real Self Views is that whereas Frankfurt is concerned primarily with *identification* at the second-order level, Watson is concerned with whether an action results from one's valuational system.

⁶ Or, as she sometimes calls it, acting in accordance with the True and The Good.

⁷ We will see, below, that Wallace has a similar view about the rational capacities relevant to responsible action being a dual power.

⁸ What does it mean to say that one "has access to the fact that setting another's house on fire is wrong." Presumably, it means that one is (a) aware that causing unnecessary harm to another is wrong; and (b) setting another's house on fire causes unnecessary suffering to another.

⁹ I will speak more about excusing conditions in section 2.3 below. My discussion, however, will not be concerned with the nature and function of *legal* excusing conditions. I believe that excusing conditions play a different role in the courtroom than they do in the moral community in which we actually live.

fire is wrong—I may, for instance, be unable to see suffering as a moral concern, or I may be unable to see how setting my neighbor’s house on fire will cause her to suffer—then it may seem inappropriate to hold me responsible for doing so. (However, if these facts are true about me, then there is clearly something wrong with me!)

It would seem similarly inappropriate to hold me responsible if I *do* recognize that setting fire to my neighbor’s house is wrong, but I am not *able* to be moved by this fact into refraining from so acting. I may, for instance, recognize that setting fire to another’s house causes unnecessary suffering and that this makes it immoral to do so, but such considerations simply have no motivational efficacy. If I am not the sort of agent who can be moved by moral considerations, how can I be a morally responsible agent?

An interesting and important consequence of Wolf’s Reason View is that moral responsibility is *asymmetrical* in nature. Because morally responsible action requires simply the ability to act according to reason, this can be satisfied in deterministic scenarios where the agent acts according to the True and the Good. Wolf offers an example of a parent jumping into a pool to save his or her drowning child. Such an individual, says Wolf, may have been psychologically determined to save the child—the parent likely did not even think about what should be done, but simply reacted to the fact that the child was drowning and the thought that jumping into the pool to save the child is what needed to be done. However, in cases where the agent fails to act according to reason—i.e. where the agent fails to act according to the True and the Good—things are different. Because morally responsible action requires the *ability* to act according to

reason, it must be the case that the agent could have done otherwise, for if she couldn't have, then the agent did not possess the ability in question.

What this asymmetry amounts to is this: only blameworthiness, but not praiseworthiness, requires the ability to do otherwise. If we are determined to do the right thing for the right reason, then in so far as we *did* do the right thing for the right reason, conditions sufficient for morally responsible action obtain. If, however, we are determined *not* to do the right thing for the right reason, then the conditions sufficient for morally responsible action fail to obtain.

Because of this asymmetry we can say that Wolf's Reason View meets the Frankfurt-style compatibilist and the libertarian halfway. Praiseworthy actions, for Wolf, are compatible with determinism in the sense that they do not require the ability to do otherwise; blameworthy actions, however, are not compatible with determinism—in order for an agent to be blameworthy for an action, it must be the case that she *could have* done the right thing for the right reason.

For the time being, the important thing to get out of Wolf's theory of morally responsible action is not her asymmetry thesis. Rather, the point I want to emphasize is Wolf's focus on the relationship between reason and action as crucial to understanding the nature of moral responsibility. For Wolf, it is in virtue of our ability to act according to moral reasons—a rational capacity of a certain sort—that makes us morally responsible agents. We can possess the ability to do otherwise, and we can identify with our actions at a second-order level, but unless we possess the ability to act according to moral reasons we cannot be morally responsible agents.

2.3 R. Jay Wallace and Strawsonian Compatibilism

Susan Wolf was, of course, not the first to emphasize the importance of rational capacities to the concept of moral responsibility. In his widely read essay “Freedom and Resentment,” Peter Strawson sketches an account of moral responsibility that is, in a sense, constructed out of our reactive attitudes. An important part of Strawson’s theory is his account of the excusing conditions. One type of excusing condition—what R. Jay Wallace will later call an *exempting condition*¹⁰—concerns the psychological state of the potential blamee. Strawson suggests that individuals who suffer from psychological abnormalities, such as psychopathy, are not appropriate targets of our reactive attitudes, and thus not blameworthy for their actions.

In this section I will look at “Freedom and Resentment” and the Strawsonian-style compatibilism it gave rise to. I will then turn my attention to the most developed Strawsonian theory of moral responsibility, namely that developed by R. Jay Wallace in his monograph *Responsibility and the Moral Sentiments*. I will use Wallace’s theory as the starting point for the skeptical argument I advance in sections 3 and 4.

2.3.1 Strawsonian Compatibilism

In “Freedom and Resentment,” Strawson takes aim at those who believe that were determinism to be true, certain practices, such as blaming and punishing, would be unjust, and related concepts, such as freedom and responsibility, would have no

¹⁰ Wallace gets the term from J.L. Austin.

application.¹¹ What is unique about the approach Strawson takes in his essay is that, rather than focus his discussion on the (un)importance of the ability to do otherwise in ascriptions of moral responsibility, Strawson instead looks at what he calls our *reactive attitudes* and their relationship to our notions of praise and blame. The question Strawson seeks to answer is the following: what effect on our reactive attitudes, and thus our practice of praising and blaming, would discovering the truth of determinism have? Interestingly, Strawson argues that (a) it would not affect our reactive attitudes and the norms that govern them; (b) it *could not* affect our reactive attitudes; and (c) even *if* it could have such an affect, it *shouldn't*. This position has become known as *Strawsonian compatibilism*.

First, a word or two needs to be said about the *reactive attitudes*. It should be noted that Strawson does not offer anything like a complete account of the reactive attitudes in “Freedom and Resentment.” Rather, Strawson offers only a sketch of what such an account might look like. One can think of the reactive attitudes as those things that link together members of a moral community. When I act in a certain way—insofar as you and I are members of the same moral community—a certain reaction on your part will be appropriate.¹² If I, for example, spurn an invitation to your birthday party citing the supposed fact that I have my thesis to write, and you catch sight of me whooping it up at a bar, it may be appropriate for you to react to this act of mine with resentment or a

¹¹ Rather than use the silly moniker that Strawson uses for these theorists, “pessimists”, I will use the more widely accepted “incompatibilist”.

¹² Which is prior, the reactive attitude relationship, or membership in the moral community? Is being a member of the same moral community what *makes it* appropriate for me to react to your actions in a certain way, or does it being appropriate for me to react to your actions in a certain way what make you and I members of the same moral community? As we will see, a question similar to this is at the heart of Wallace’s investigations.

similar attitude that conveys the fact that I have violated a norm. If, on the other hand, after inviting me to your party I show up with a thoughtful gift, the appropriate attitude to take toward me will not be one of resentment, but rather gratitude. The idea is that as members of the same moral community certain attitudes you take toward my actions will “fit” appropriately, while others will not. Your resentment would not fit my gift-giving, nor would your gratitude fit my invitation-spurning. That we take certain attitudes to be appropriate candidates as reactions to certain kinds of actions and inappropriate to other kinds of reactions is (often seen as) evidence that these reactive attitudes are *norm-governed*.

For Strawson, resentment and gratitude are only two of the many reactive attitudes. Other attitudes he considers are indignation, guilt (a self-directed reactive attitude), forgiveness, anger, and love (among others). Like Strawson, I will focus my discussion on resentment and indignation. Doing so will make it easier to focus in on the issue of moral responsibility and, for the time being, its relationship to determinism.¹³

Let’s assume for the sake of discussion that the case has been made that the reactive attitudes are in fact governed by norms. Presumably, these norms dictate, first, which attitudes correspond to which actions, and second, under what circumstances the attitudes are appropriate responses to those actions. As it relates to the reactive attitude relationship, the position of incompatibilism can be understood as maintaining that, if determinism is true, then certain (maybe even all) reactive attitude relationships would

¹³ I will be jettisoning my concern with determinism in the sections that follow.

not be appropriate. In other words, for the incompatibilist, certain reactive attitude relationships are only appropriate in indeterministic settings. The incompatibilist will argue that resentment, indignation, and guilt¹⁴ will only “fit” an action when the agent performing that action *could have avoided* performing it (and one can’t avoid performing an action in deterministic settings). Just as Real Self compatibilism holds that the ability to do otherwise is irrelevant to matters of moral responsibility, so does Strawsonian compatibilism hold that the ability to do otherwise is irrelevant to the reactive attitude relationship.¹⁵

The first step of Strawson’s argument for rebutting the incompatibilist involves examining those cases where we *do* hold that the reactive attitude relationship is inappropriate (that is, cases where it is commonly deemed inappropriate to react to another’s action with, e.g., resentment) and then argue that the structure of these cases is such that they cannot serve as an analog to determinism.

The reactive attitude relationship Strawson considers is that of *resentment*.¹⁶ When we react to another with resentment, it is generally as a result of some harm—either emotional or physical—that has been brought upon us by another. Sometimes, however, we will learn certain facts about the individual wronging us that *mitigates* our reaction. Strawson distinguishes between two broad groups of such factors. The first group—what I will call the type-1 group—contains those circumstances that point to the

¹⁴ As we shall see, Wallace takes these three reactive attitudes to be the ones that hang together as class to form the concept of moral responsibility.

¹⁵ It should be noted that, technically, Strawson does not connect the reactive attitude relationship to moral responsibility *per se* (though there are remarks that suggest that he takes there to be such a connection). The connection between reactive attitudes and moral responsibility is made perspicuous by Wallace.

¹⁶ The “vicarious analog” to resentment is *indignation*. We react with indignation, not when one harms us, but when one harms *another*.

fact that the harm being brought upon the victim was not a result of *actual* wrongdoing. It may have been that the perpetrator acted under excusable ignorance or that the harm the perpetrator brought upon me resulted from an unfortunate (non-negligent) accident. In such cases, because the agent was acting in good faith—she or he did not have ignoble intentions—it is generally deemed that resenting him or her for the act is inappropriate.¹⁷

The second group of mitigating factors concern features of the agent—rather than the circumstances of the action itself—that make resentment an inappropriate reaction. This group of mitigating factors, which Strawson in turn divides into two sub-groups, concerns what we might call the “psychological make-up” of the transgressing agent. The first of these subgroups—I will call this the type 2a group—concerns those circumstances where the transgressing agent is *temporarily* incapacitated, as when one endures great duress (a paradigm case being soldiers in combat situations¹⁸) or when one is being emotionally manipulated (a limiting case being action resulting from post-hypnotic suggestion). The second of these subgroups—what I will call the type-2b group—concerns features of the transgressing agent that *permanently* incapacitate him or her. Examples of such features include certain mental and mood disorders, such as

¹⁷ Wallace has an excellent discussion of this in *Responsibility and the Moral Sentiments*, 120-127. For Wallace, in such cases the reason that blame is mitigated is that *wrongdoing* requires intentional action on the part of the offender. If, in acting, the agent is not acting intentionally, then the agent is not, in fact, doing wrong. And since wrong-doing is a necessary condition for blameworthiness, blame is not appropriate in such cases. Wallace’s account of the excuses and exemptions will be discussed in 2.3.2 and play an important role in the argument of section 4.

¹⁸ For an article that defends the non-culpability of certain war-time atrocities see John M. Doris and Dominic Murphy, “From My Lai to Abu Ghraib: The Moral Psychology of Atrocity,” *Midwest Studies in Philosophy*, XXVI (2007): 25-55.

schizophrenia and severe depression, and compulsive behavioral disorders, such as kleptomania and pyromania.¹⁹

Strawson suggests that when an agent falls within the second group of conditions we are invited to:

suspend our ordinary reactive attitudes towards the agent, either at the time of action [in the case of the type-2a group] or all the time [in the case of the type-2b group]. [The mitigating factors] do not invite us to see the agent's action in a way consistent with the full retention of ordinary interpersonal attitudes and merely inconsistent with one particular attitude. They invite us to see the agent *himself* in a different light from the light in which we should normally view one who has acted as he has acted.²⁰

When an agent harms us accidentally, we still see the agent, all else being equal, as an appropriate "target" of our reactive attitudes. In suspending our reactive attitudes in such circumstances, we are merely recognizing that the event in question (that which has brought harm to us) is not as we had first viewed (e.g. as an intentional violation of our rights). On the other hand, when we suspend our reactive attitudes toward an agent under type-2 conditions, when we see that the agent suffers from some psychological disorder or some cognitive impairment, we *do* alter our overall attitude with respect to the agent.

¹⁹ Interestingly, Strawson includes *being a child* in the type-2b group of mitigating features. In a certain respect, I think this is a correct. When a child harms us we generally find it inappropriate to react to his or her actions in the way we would were the same harm brought upon us by an adult. Of course being a child is not a permanent feature of an individual. This gives reason to think that that the groups should not be divided in terms of permanence/impermanence but rather in terms of *duration* (e.g. long-term/short-term). If we separate along the latter lines, however, we run into something of a Sorites paradox: where do we draw the line between long and short? In the end, I don't think the way in which we make the cut-off matters all that much. The designation of subgroup is merely a heuristic device to think about the type-2 group itself, which *is* sufficiently different in kind from the type-1 group to set it apart conceptually.

²⁰ Strawson, "Freedom and Resentment", p. 78.

In suspending our reactive attitudes in this way, we adopt an overall attitude Strawson calls “the objective attitude,”²¹ an overall attitude wholly different from—and incompatible with—the attitudes necessary for our ordinary interpersonal relationships.²²

According to Strawson, when we adopt the objective attitude toward another, we see him or her “as an object of social policy; as a subject of what, in a wide range of sense, might be called treatment; as something certainly to be taken account, perhaps precautionary account, of.”²³ This overall attitude is quite different from what Strawson calls “the participant attitude,” (the attitude of ordinary interpersonal relationships). In the participant attitude, we view the other agent, not as an object of social policy, but as a fellow member of our moral community.

We can now turn to Strawson’s three arguments for compatibilism. The first argument we can call “the similar-structure argument.” According to the first premise of this argument, if determinism is incompatible with the reactive attitude relationship, then we would expect to find something similar to the structure of determinism in the excusing conditions outlined above. Strawson rules out the factors appealed to in the type-1 scenario—such as ignorance or accident—as irrelevant to incompatibilism. The truth of determinism will change nothing about my underlying intentions in action in that it would not alter the relation between my intention and action.²⁴ With regards to the

²¹ Strawson, “Freedom and Resentment,” p. 79.

²² I use the phrase “overall attitude” in order to distinguish a single attitude—say *resentment*—from the conglomeration of attitudes that make up any perspective. It is this conglomeration that ‘overall attitude’ refers.

²³ Strawson, “Freedom and Resentment,” *ibid.*

²⁴ Interestingly, among the “pleas” Strawson uses to capture the type-1 excusing condition are “he couldn’t help it” and “he had to do it” (“Freedom and Resentment,” p. 77). These *do* seem remarkably similar to the structure of determinism.

type-2 excusing conditions, Strawson writes, “it cannot be the consequence of any thesis which itself is not self-contradictory that abnormality is the universal condition.”²⁵ We adopt the objective attitude under type-2 conditions insofar as the agent who has harmed us suffers from abnormal psychological “defects” and thus is not capable of maintaining normal human interpersonal relationships. I take it that Strawson’s suggestion is not just that it would be an absurdity to say that abnormality is the norm, but also that determinism would not, in itself, make us incapable of maintaining relationships in the way that morality requires. It would seem, then, that determinism would yield us neither type-1 nor type-2 excusing conditions.

The way in which Strawson presents this will certainly not convince those committed to the principle of alternate possibilities. To such individuals, Strawson presents two arguments. According to the first argument—what Wallace calls “the naturalistic argument”²⁶—universally adopting the objective attitude is simply something that humans cannot do. Even if it were true that determinism would universally excuse, adopting the corresponding attitude—viewing ourselves and others as objects of social policy rather than persons—would result in the destruction of interpersonal relationships as we know it, for interpersonal relationships *require* the reactive attitudes. But interpersonal relationships are an essential and ineliminable part of our lives (or what we might call “the human condition”). Because of this, even if we were to find out tomorrow that determinism is true and that everyone is universally

²⁵ “Freedom and Resentment,” p. 81.

²⁶ This moniker comes from Wallace. cf. *Responsibility and the Moral Sentiments*, p. 96.

excused, shedding the reactive attitudes and adopting the objective attitude is simply something that we cannot do.

This *naturalistic argument* will appear to many to be more of an anthropological (and hence empirical) claim than a philosophical claim. We simply cannot assess the truth of such a claim from our armchair. Strawson's next argument—what Wallace calls the “pragmatic argument”—can be understood as a reply to this objection.²⁷ According to the pragmatic argument, even if we *could* universally adopt the objective attitude, not only would we not *want* to, we in fact *shouldn't*. The argument for this claim is somewhat similar to the naturalistic argument. Interpersonal relationships—to which, again, the reactive attitudes are essential—are an *important* part of our lives. To universally adopt the objective attitude would require abandoning these relationships. But the payoff is just not worth the cost. Even if, in fact, determinism universally excuses, we should act as if it did not. To do otherwise would be to spurn the most important part of our lives.

Of the three arguments, it is Strawson's first and third that get the most sympathy. As we shall see shortly, Wallace's theory of responsibility is basically a refinement of the first argument.

2.3.2 Wallace and Strawsonian Compatibilism

It may not be immediately clear why I've included Strawsonian compatibilism under the heading of rational capacity compatibilism, for it may appear that Strawson's

²⁷ Cf. Wallace, *Responsibility and the Moral Sentiments*, p. 96.

primary concern is not with the role that our rational capacities play in ascriptions of blame and responsibility, but rather with the role of our *sentiments* or *reactive attitudes* in such ascriptions. The reactive attitudes, however, are not the only player in Strawson's theory. There is an important element that holds all the pieces together. This is the *normative* component. And it is this normative component that makes Strawsonian compatibilism fall under the heading of Rational Capacity Compatibilism.

The normative component comes in to play when we start considering when and whether an agent is an *appropriate* target of our reactive attitudes. In Strawson's essay, we find the discussion about norms buried in the talk of "excusing conditions." Strawson's type-1 excusing condition answers the question of *when* an agent is an appropriate target (namely when her actions are *in fact* moral transgressions). The type-2 excusing conditions answer the question of *whether* an agent is an appropriate target of our attitudes. It is here that we find a concern with the rational capacities. And while this concern is not exactly perspicuous in Strawson's essay, we find it more clearly spelled out in the account R. Jay Wallace provides in his monograph *Responsibility and the Moral Sentiments*, which owes a heavy debt to Strawson. With the aim of fleshing out the relationship between the reactive attitudes, the rational capacities, and the practice of blame, I turn now to Wallace's brand of Strawsonian compatibilism.

2.3.2.1 *A Theory of Responsibility out of an Account of the Reactive Attitudes*

One of the major contributions of Wallace's work is his construction of a robust account of moral responsibility out of the norms governing our reactive attitudes.²⁸

Strawson's argument can be understood as going through several stages. The first stage consists of an analysis of the moral emotions, or reactive attitudes. The second stage consists in an analysis of the concept of holding responsible. The final stage consists in an account of the norms that govern the practice of holding responsible. It is in this, the final stage, that we unearth the role of the rational capacities in the practice of blame. In this subsection (2.3.2.1) I will go through the first two stages of Wallace's argument. In the next subsection (2.3.2.2) I will go through the third stage.

In his analysis of the moral emotions, Wallace concerns himself with two questions: (a) what are the "essential features" of the moral emotions; and (b) what "holds [the moral emotions] together as a class?"²⁹ Wallace considers the following to be essential features of the moral emotions: (i) to have a moral emotion is to react in a certain way with a certain attitude, that is to say, to experience a reactive attitude; (ii) to have a moral emotion is to take part in "a distinctive form of evaluation," a form of evaluation that Wallace calls "holding a person to an expectation."³⁰ While Wallace thinks that (i) and (ii) are constitutive of the moral emotions, he thinks that they cannot be understood in isolation from another. That is, he doesn't think that one is conceptually prior to the other.

²⁸ Admittedly, his account is fairly one-sided in that he only accounts for the negative aspect of moral responsibility, that side which we associate with blame.

²⁹ Wallace, *Responsibility and the Moral Emotions*, p. 18

³⁰ *Ibid*, p.19

Following Wallace, we can distinguish between two kinds of expectation.³¹ The first is what we might call an *epistemic* expectation.³² To have an expectation of this kind is to have a belief that there is a high probability that a certain even will occur. Among my most reasonable expectations are that the sun will continue to rise, the stars will continue to set, and the Padres will fail to make the World Series yet again this season. While these expectations can be associated with emotions—I often get frustrated by the Padres’ constant poor performance—epistemic expectations are not, in general, tied to any distinctively *moral* emotion.

The second kind of expectation—what I will call a *normative expectation*—is tied to particular emotions that *are* distinctively moral. This kind of expectation is one that we *hold* people to. My expectation that, for example, the Padres will not cheat by taking steroids is unlike my expectation that the Padres will fail to make the World Series. The way I will react if the former expectation is not met will differ from the way I will react if the latter is not met. Were I to find out that my favorite Padre was taking steroids, I would likely react with indignation, whereas if the Padres actually end up making the World Series (if only!), I would likely be pleasantly surprised. It is important to recognize that the indignation that I would feel toward a Padre upon finding out that he is in fact cheating is not (necessarily) tied to any probability. Even if I find it highly probable that the Padres are using steroids (given, for instance, the size of their biceps),

³¹ The two different kinds of expectation will arise again, and become relevant, in section 3.

³² The ‘epistemic expectation’ moniker is my invention. Wallace doesn’t give them a name.

it would still be appropriate to feel indignant toward them upon discovering the truth.

This is because the expectation in question acts more like a demand than a prediction.³³

The relationship between a normative expectation and the reactive attitudes is not just one of happenstance. Rather, when an expectation to which we hold someone is violated, we *deem it appropriate* to react with indignation, resentment, or guilt. The class of action to which the moral emotions are appropriate reactions are *constrained* by our normative expectations. But our normative expectations are understood in virtue of the fact that moral emotions are the appropriate kind of responses to such violations.

Wallace recognizes that Strawson identifies three different kinds of reactive attitudes: *personal reactive attitudes*, such as resentment, love, and forgiveness; *moral reactive attitudes*, such as indignation and disapprobation; and *self-reactive attitudes*, such as guilt, remorse, and shame. What is distinctive about personal reactive attitudes, claims Strawson, is that they are the “non detached”³⁴ attitudes involved when one person interacts directly with another, where the one who reacts demands “a degree of consideration toward oneself.”³⁵ Strawson understands moral reactive attitudes to be “vicarious analogues” of the personal attitudes, in so far as the standards that one is being held need not be in relation to oneself (e.g. when I respond to Moammar Gadhafi with indignation, the obligation which I take him to be violating is not one that relates to me—it relates to his citizens). Finally, Strawson understands self-reactive attitudes as

³³ And thus why I characterize it as *normative*.

³⁴ Strawson, “Freedom and Resentment”, p. 62; quoted in Wallace, *Responsibility and the Moral Sentiments*, p. 35.

³⁵ Wallace, *Responsibility and the Moral Sentiments*, p. 35.

those attitudes related to those demands we make on ourselves when we interact with others.

Wallace takes issue with Strawson's understanding of the moral reactive attitudes. He proposes to understand the difference between a moral and non-moral emotion in terms of the kind of expectation. Ideally, when we hold another to an expectation, we are justified in doing so, and to be justified in doing something is to have reasons that support doing that thing.³⁶ When the reasons that support an expectation are moral reasons, Wallace wants to call the expectation in question a *moral expectation*, or what is the same (to him), an *obligation*. It is the connection with an obligation that, for Wallace, designates a reactive attitude as moral. Understanding the moral reactive attitudes in this way, "narrows the class" of attitudes to three: *resentment*, *indignation*, and *guilt*. It will be noticed that only the second of these attitudes was identified by Strawson as distinctly moral. Resentment was viewed as a non-moral personal-reactive attitude. Under Wallace it becomes moral in so far as resentment is an appropriate response to another who violates an obligation that he or she has towards you. Likewise, *guilt* can be understood as a moral emotion in that it is an appropriate response to a self-violation of an obligation.³⁷

With this outline of the moral emotions in hand, we are now in a position to turn to Wallace's analysis of what it means to be held morally responsible. We can begin

³⁶ Of course, we can unreasonably hold someone to an expectation, as when my mother unreasonably holds me to the expectation that I become a lawyer like her.

³⁷ Saying either of these is not to discount the possibility of *irrational* emotions, such as irrational guilt (as when I feel guilty for breaking up with my partner even though I was fully justified in doing so) or irrational resentment (as when I resent my successful colleague not because she has violated an expectation, but simply because she is successful in a way that I am not).

with the familiar distinction that most theorists make when articulating such a theory: that between *moral* and *merely causal* responsibility. Whatever account we offer of moral responsibility, it must be able to tell us what distinguishes it from the way that the Moon is responsible for the ocean's tides. To say that the moon is responsible for the tide is to say nothing other than that the moon *caused* the tides. But when we say that, for example, Bernie Madoff is morally responsible for running a Ponzi scheme, we mean more than just that he was the just cause or source of the Ponzi scheme.³⁸ It is the job of a proper theory of moral responsibility to say what needs to be added to the notion of responsibility in the merely causal sense to produce the notion of responsibility in the robust moral sense.

One thing that sometimes gets said³⁹ is that what distinguishes moral responsibility from merely causal responsibility is that the former is *deep* in a way that the latter is not. Wallace is quick to point out, however, that while depth may be a necessary component of moral responsibility, there is a third kind of responsibility that satisfies the "depth" requirement but is distinct from responsibility in the moral sense. Wallace calls this third kind of responsibility "the condition of autonomy."⁴⁰ It is this notion of responsibility that is in play when we credit or praise an artist for a magnificent work of art. In praising the artists, we are not simply recognizing the fact that the artist happened to be the cause of the work of art. "Rather," Wallace claims, "our praise and admiration reflect a kind of credit on [the work of art's] creator."⁴¹ What is more, "we

³⁸ Though being causally responsible may certainly be a necessary condition for moral responsibility.

³⁹ Wallace references Susan Wolf's *Freedom within Reason* here.

⁴⁰ Wallace, *Responsibility and the Moral Sentiments*, p. 52.

⁴¹ *Ibid*, p. 52-53.

can say that the artist is responsible for the work of art in a way that we cannot say that a very young child is responsible for her finger paintings.”⁴² To take such an attitude toward the mature artist is to recognize, in a *deep* sort of way, that she is *responsible* for her work of art. But this is certainly not the sense of responsibility we have in mind when discussing Bernie Madoff and his Ponzi scheme.⁴³

So if moral responsibility is something over and above mere causal responsibility, and distinct from the deep kind of responsibility we find in the condition of autonomy, what is it? The move Wallace makes here is interesting: we can’t understand what it is *to be* morally responsible in isolation of what it means *to hold* someone morally responsible. Or, to be more precise, the stance of *holding* responsible is conceptually prior to the notion of *being* responsible. To understand moral responsibility, we must turn to the stance of holding responsible .

The idea here is a deep and important one. In making this claim Wallace is saying that moral responsibility is not something *wholly metaphysical*. Rather, Wallace is claiming that moral responsibility is *practice dependent*. That is, the facts about responsibility are dependent on our practices, *not independent of them*.⁴⁴ An agent can be morally responsible only in a society where the practice of holding responsible exists.

So what is it *to hold* someone morally responsible? The natural move here is to relate the moral reactive attitudes—and the stance of holding to an expectation associated with them—to the practice of holding responsible. Wallace puts it this way:

⁴² *Ibid*, p. 53.

⁴³ Though we could imagine a criminal greatly admiring the sheer audacity of Madoff’s fraud in the same way that we admire a great work of art.

⁴⁴ Cf. Wallace, *Responsibility and the Moral Sentiments*, p.89.

“to hold someone morally responsible is to hold the person to moral expectations [i.e. obligations] that one accepts.”⁴⁵ And to hold someone to moral obligations that one accepts, is to be susceptible to the moral reactive attitudes were the obligations to be violated. It is important to note that susceptibility here is a normative notion. We can unpack it thus: to be susceptible to the moral reactive attitudes is to deem it *appropriate* to react in a certain way (i.e. with resentment, indignation, or guilt) when one violates an obligation that we accept.

With a sketch in hand of what it is to *hold* someone morally responsible, we can now turn to Wallace’s account of what it is to *be* morally responsible. Wallace offers the following: “*S* is morally responsible (for action *x*) if and only if it would be appropriate to hold *S* morally responsible (for action *x*).”⁴⁶ Two things should be noted about this approach. First, the claim is not that one is responsible *when* one is held morally responsible, but rather when it is *appropriate* to hold someone responsible. This means that *actually* holding one morally responsible is neither necessary nor sufficient for one to *be* morally responsible. Second, this schema brings out how moral responsibility is governed by norms, namely those gauging when it is *appropriate* to hold someone morally responsible. Wallace’s account, then, rests on two things: (1) his account of the reactive attitudes I summarized above; and (2) an account of the *norms* governing when it is and is not appropriate to take the stance of holding someone responsible. It is through (2) that we discover the importance of the rational capacities in Strawsonian compatibilism. I turn to that next.

⁴⁵ Wallace, *Responsibility and the Moral Sentiments*, p. 65.

⁴⁶ Wallace, *Responsibility and the Moral Sentiments*, p. 91.

2.3.2.2 *The Norms at the Heart of Strawsonian Compatibilism*

Any adequate account of the norms that govern the practice of holding responsible will have to answer two questions. First, the account must answer the question of *what kinds of norms* are relevant to moral responsibility. Second the account must answer the question of when these norms are satisfied. I will go through Wallace's answer to each of these questions in turn.

To say what kinds of norms are relevant to responsibility will be to say what it means for it to be *appropriate* to hold someone morally responsible. One possible way to answer this—one that Wallace rejects—is the consequentialist's strategy, or what he calls *the economy of threats* approach. According to this approach, it is appropriate to hold another morally responsible (for an action ϕ) if doing so would be, on the whole, better than refraining from holding that person responsible (for ϕ). This usually gets cashed out in terms of motivation: if ϕ is an act we would prefer S not to perform (presumably we need to be justified in our preference) then it is appropriate to hold S responsible for ϕ if doing so would motivate S (and maybe others) to refrain from ϕ -ing. Thus, for the economy of threats approach, an agent's blameworthiness is not about the agent, but about motivational efficiency. Whatever attitude will produce the best result is the attitude that ought to be adopted.

There are many reasons to reject the economy of threats approach. Chief among them is the fact that in it we lose any sense of the practice of blame as we know it. For if we are really just concerned with deterrence, any reaction that gets the job done will do, and because more severe sorts of reactions are likely to work better as a deterrent

than milder reactions, the economy of threats approach will justify severe penalties for what we might consider to be minor offenses.⁴⁷ Wallace offers another reason to reject the economy of threats approach. One of his desiderata for an adequate theory of responsibility is that it take seriously incompatibilist intuitions regarding moral responsibility and determinism. The economy of threat approach fails in that it is blatantly a compatibilist approach.

The approach that Wallace prefers cashes out appropriateness in terms of *fairness* and *reasonableness*: it is appropriate to hold another morally responsible only if is both *fair and reasonable* to do so. When is it fair and reasonable to do so? At this point, we've hit something close to a conceptual rock bottom. We can say that fair and reasonable are moral concepts linked to the concept of desert. But saying this doesn't do much to unpack the concepts at hand. The better strategy is to take an approach similar to that of reflective equilibrium: rather than perform a strict conceptual analysis, we can reflect on our practices of holding responsible and through this reflection discern a principle that underlies the practice.

Staying true to his Strawsonian roots, Wallace considers conditions that are generally taken to excuse and exempt agents from blame. In doing so, Wallace is not only concerned with uncovering the principles that guide our practice of blame; an ulterior motive of his is to strengthen Strawson's "similar structure argument."

⁴⁷ This, of course, is an empirical question. But even if it turns out that more severe forms of punishment don't in *actuality* work as a deterrent, we can imagine that on some possible world they do, and on this world these severe forms of punishment would be justified by the economy of threats approach.

Like Strawson, Wallace divides these conditions into two classes. The first class he calls *excusing conditions* (this class roughly corresponds to Strawson's type-1 group); the second class he calls *exempting conditions* (this class roughly corresponds to the type-2 group). We can draw the line between an excuse and exemption by noting that the excuses function *locally* (i.e. on single events) while exemptions operate *globally* (i.e. they [can] range over several events).⁴⁸ Another way to make this distinction is to say that the excuses excuse *actions*, while exemptions exempt *agents*.

The first thing to note when thinking about excusing and exempting conditions is that they arise for what I will call *prima facie* moral transgressions, that is, actions that appear on the face of it to violate a moral obligation. Now, we can say that what separates a justification for ϕ -ing from an excuse for ϕ -ing is that a justification shows that ϕ is not in fact morally wrong, whereas an *excuse* "grants that [ϕ] would [normally] be wrong." but since the action was done unintentionally, no moral obligation has in fact been violated.⁴⁹

A difficult question arises when we ask why unintentionally ϕ -ing gets one off the hook whereas one would be on the hook for ϕ -ing were one to ϕ intentionally. Wallace's answer to this is instructive and goes a long ways to uncovering the normative principles underlying moral responsibility. As I stated above, Wallace understands the stance of holding another responsible in terms of holding another to an obligation that you yourself accept. Now, in order to hold another to an obligation, that person must be able to act in accordance with this obligation. But to act in accordance

⁴⁸ Wallace, *Responsibility and the Moral Sentiments*, p. 118.

⁴⁹ Wallace, *Responsibility and the Moral Setniments*, p. 119.

with an obligation is to act in accordance with the reasons that support that obligation. When we morally assess an individual's actions, we are not assessing the bodily movements themselves, so much as we are assessing *the quality of the person's choice in conforming or failing to conform with an obligation that we accept*. That is, to assess an agent's action is to assess the reasons the agent chose to act on. But if a bodily movement is not intentional, then it is not a product of a reason for action, and thus not reflective of any choice that the agent has made. Though the unintentional bodily movement may have the appearance of violating a moral obligation, insofar as it was not the product of a reason for acting, no such violation could have in fact taken place.

To *excuse* an agent for ϕ -ing, then, is to judge that in ϕ -ing the agent did not in fact violate an obligation. To blame an agent for an act that does not violate an obligation is unfair in that "it is surely the case that people do not deserve to be blamed if they have not done anything wrong in the first place."⁵⁰ Wallace calls this "fundamental principle of desert" the "principle of no blameworthiness without fault."⁵¹

Whereas conditions that excuse an agent for an act focus on features of the act that make it "unfit" (so to speak) for blame, conditions that *exempt* an agent focus on features of *the agent* that make her an "unfit" target of blame. Like Strawson, Wallace separates exempting conditions into two subgroups. The first subgroup concerns those conditions that *temporarily* (or locally) render an agent unfit, while the second subgroup renders the agent unfit under normal conditions (or globally). Because the difference

⁵⁰ Wallace, *Responsibility and the Moral Sentiment*, p. 135.

⁵¹ *ibid.*

between these two groups is only a difference of duration, and not a difference in kind, I will treat them as a homogenous group.

The guiding question underlying the exemptions is the following: in virtue of what does an agent become an unfit target for blame? As it came out in the discussion of the excusing conditions, when we hold an agent morally responsible, the object of our moral assessment is the quality of the agents' choice in regards to her moral obligations. Obligations, which are supported by reasons for acting, are violated when an agent disregards these reasons and acts contrary to them. But in order for the agent to be open to moral assessment in the first place, she must be *able* to act in accordance with her moral obligations, and this requires the possession of certain rational powers, or what Wallace calls *the powers of reflective-self control*.

Wallace distinguishes between two sets of rational powers. The first is "the power to grasp and apply moral reasons;" the second is "the power to control or regulate [one's] behavior by the light of such reasons."⁵² I will discuss each of these powers in turn.

The power to grasp and apply moral reasons is a bipartite power. The first, the power to grasp moral reasons, involves the ability to understand the concepts employed in moral principles. Wallace offers as an example the principle of nonmalficence: commit as little harm as possible in the pursuit of one's ends. A person will be able to abide by this principle only if she understands the concept HARM. Truly grasping the concept involves more than merely knowing how to correctly use. It involves, among

⁵² Wallace, *Responsibility and the Moral Sentiments*, p. 157.

other things, knowing what makes HARM a morally important concept. But simply understanding the concept HARM in this way is not enough. One must also be able to *apply* the concept in practice—this is the second bipartite power, which involves, for example, being able to distinguish permissible from impermissible harm, being able to predict whether one’s actions are likely to cause harm, and being able to perceive when others are in fact harmed.

The second power—the power to regulate and control one’s behavior by the light of moral reasons—has three component powers: (1) the ability to critically reflect on and assess one’s reasons; (2) the ability to deliberate and make choices as a result of deliberation; and (3) the power to translate choices into behavior. An agent lacking the first component power will be similar in many ways to the wanton featured in Harry Frankfurt’s writings who simply acts on whichever impulse happens to be the strongest at the time. An agent lacking the second power is one who lacks the ability to reason practically; she will be unable to take reasons as premises in a practical syllogism and foresee those reasons into a choice. Finally, an agent lacking the third power will be unable to translate her choice into *action*. We often think of the unwilling heroin addict as lacking this power when she is caught in the grips of a craving that compels her to score more dope.

According to Wallace, when an agent lacks one of the above powers, she is exempted from blame. Wallace reasons that to blame an agent when she lacks one of the above powers would be unreasonable, for it would require holding her to an obligation that she either lacks the ability to understand or comply with.

Given that the so-called powers of reflective self-control are among one's rational capacities, it should now be evident as to why Strawsonian style compatibilism falls under the heading of rational capacity compatibilism. While Strawsonian compatibilists are concerned with the reactive attitudes in a way that, for example, Susan Wolf and others are not, Strawsonian compatibilists also give a central role to one's rational capacities when it comes to ascriptions of moral responsibility.

2.3.2.3 Wallace's Version of the Similar-Structure Argument

As discussed in section 2.3.1, Strawson argues, contra the incompatibilist, that since the exempting conditions are not similar in structure to the thesis of determinism, there is no *prima facie* reason to think that determinism is incompatible with blameworthiness. There is, however, a retort that Strawson does not consider. According to the exempting conditions (or type-2 excusing conditions), an agent is exempted from blame if she lacks one of the relevant powers of reflective self-control. According to Wallace, "people will succeed in exercising these powers only when they actually comply with moral obligations and do so because they grasp the reasons that support them."⁵³ At this point the incompatibilist is likely to interject: in deterministic settings, if at *t* agent *S* fails to fulfill her obligation, then at *t* she is not *able* to exercise her powers of reflective self-control; and if she is not *able* to exercise said powers, then it would be unfair to blame her for her action at *t*.

⁵³ Wallace, *Responsibility and the Moral Sentiments*, p. 190.

Wallace's reply to the incompatibilist is somewhat unsatisfying: "what matters is not the exercise of the general powers of reflective self-control, but the possession of such powers."⁵⁴ It's not clear to me why mere *possession* of the capacities warrants blameworthiness even in situations where one cannot gain access, so to speak, to the capacities. Consider the following scenario:

Trust Fund: Bernie's grandparents left him a trust fund of \$2 million dollars.

However, as per the rules of the will, Bernie is not allowed to touch the trust fund until he is \$60 years old, *for any reason*. Bernie is now 25 and has been out of work for three months and is about to lose his apartment if he doesn't come up with rent.

In **Trust Fund**, Bernie has a trust fund such that, were he to access it, he would be able to keep his apartment. Of course, Bernie *can't* access the trust fund and so *can't* pay off his apartment. But because Bernie has possesses a trust fund, are we to *expect* that he pay off his rent?

The case of **Trust Fund** seems analogous to cases where agent's fail to exercise their rational capacities in deterministic scenarios. In such scenarios they have possession of rational capacities that they simply can't exercise. But just as it seems unreasonable to expect Bernie to pay off his rent using his trust fund, I submit that it is unreasonable to hold agents to an obligation in deterministic scenarios where they have failed to discharge those obligations, for just as Bernie is unable to access the funds

⁵⁴ Wallace, *Responsibility and the Moral Sentiments*, p. 190.

required to pay off his rent, determined agents (who have gone wrong) are unable to exercise their rational capacities in deterministic scenarios.

I don't believe that this line of argument will likely convince Wallace.

Ultimately, however, I am not concerned with the issue of the compatibility of moral responsibility with determinism. I raised this argument here mainly because we will see the strategy Wallace uses to dispose of the incompatibilist sprout up again in section 4 as a possible objection to my argument. For the rest of this essay, I set aside the issue of determinism.

3. THE SKEPTICAL ARGUMENT I: ROSEN'S SOFT SKEPTICISM

This section marks the first step of my skeptical argument. This step consists primarily in defending a skeptical argument advanced by Gideon Rosen in his paper, “Skepticism about Moral Responsibility.” The substantive thesis of Rosen’s argument concerns the relationship between *akrasia*, ignorance, and moral responsibility. The relationship that stands between these three notions is this: an agent can be responsible for an act only if *akrasia* falls somewhere in the act’s etiology; if *akrasia* is not found within an act’s etiology—if the act bottoms out in ignorance—then the agent cannot be responsible for said act.

The layout of this section is as follow. In 3.1, I introduce the notion of *akrasia*, briefly discussing how I will be using the term. In 3.2, I offer a brief exegesis of Rosen’s argument. In 3.3 I discuss a recent criticism of Rosen’s argument advanced by William FitzPatrick. Finally, in 3.4 I defend Rosen, making use of a recent argument by Neil Levy.

3.1 *Akrasia*: What It Is

I begin by getting straight on terminology. I shall be using the term ‘*akrasia*’ in a technical sense, to denote a very specific kind of action. The kind of action I wish to denote with the term is *intentional action done against one’s best judgment*. There is quite a bit of literature on the issue *akrasia*, and it has become quite widespread to use the term ‘*akrasia*’ interchangeably with the phrase ‘weakness of will.’ However, like

Richard Holton,⁵⁵ I believe this practice has led to confusion. As Holton notes, the folk concept of weakness of will is quite different from the philosopher's concept of *akrasia*; it more closely corresponds to a judgment shift in the face of temptation than to an intentional action against one's best judgment. In getting a picture of how these notions differ, it might help to define them in a more formal manner:

Akratic action: In ϕ -ing, S acts akratically at t if and only if at t , S judges that (all things considered) she ought not ϕ but ϕ s anyway, *without her judgment changing*.

Weakness of Will: In ϕ -ing, S acts from weakness of will at t if and only if at t_n S judges that (all things considered) she ought not ϕ , but at t_{n+m} S 's judgment shifts such that she no longer thinks that (all things considered) she ought not ϕ and as a result S ϕ s at t_{n+m} .

Ultimately, no part of my argument hangs on my terminology (though it should be noted that Rosen is in agreement with Holton on this issue).⁵⁶

3.2 Rosen's Argument

We can distinguish between two kinds of skepticism about blameworthiness—a hard or severe form of skepticism, and a moderate or soft form of skepticism. According to the hard form of skepticism, blameworthiness is an impossibility—no one ever is or can be blameworthy for anything. The soft skeptic, on the other hand, holds that agents *can* be blameworthy and may in fact *be* blameworthy in many cases. The soft skeptic,

⁵⁵ Holton, Richard, *Willing, Wanting, Waiting* (Oxford: Oxford University Press, 2009).

⁵⁶ As it happens, it will also help to understand a move in Gideon Rosen's soft skeptical upshot.

however, believes that given certain contingent features of the world and our own psychology, we can never be *certain* whether any given individual at any given time is blameworthy for any given act. Because of this, the soft skeptic advocates a suspension of judgment with regards to blameworthiness.

The argument that I advance in section 4 is a hard skeptical position with regards to blameworthiness. I argue that, given our best conception of blameworthiness and our understanding of excusing and exempting conditions, no one can ever satisfy the conditions for blame (at least as they are set out by R. Jay Wallace). In this section, however, I look at a recent soft skeptical position advanced by Gideon Rosen,⁵⁷ according to which “confident positive judgments of blameworthiness [can] never [be] justified.”⁵⁸ My interest in Rosen’s argument, however, lies not in his soft skeptical conclusion but in an important lemma of his argument. This lemma—which is a substantive philosophical position in itself—will play an important role in the hard skeptical conclusion I advance in section 4. My aim in this section is to present Rosen’s argument for this lemma, and discuss the conclusions he reaches with it. In 3.3-3.4 I will defend this lemma.

⁵⁷ Gideon Rosen, “Skepticism about Moral Responsibility,” *Philosophical Perspectives* 18 (2004): 295-313.

⁵⁸ *Ibid.*, p. 296.

3.2.1 *Derivative and Original Responsibility*

The first step of Rosen's argument draws an intuitive distinction between what Rosen calls "original" and "derivative" responsibility.⁵⁹ In many cases of *prima facie* blameworthy action, an agent is responsible simply in virtue of committing the act in question. Imagine I walk up to you and slug you in face. Assuming no excusing or exempting conditions obtain, the blame that I deserve for this act is deserved simply in virtue of the fact that (a) I have the relevant cognitive capacities; and (b) my slugging you in the face violated a norm that you (and I) ought to accept. In such instances, my responsibility is *original*—the blame I receive (or ought to receive) is traced only to the original transgression. It might be, however, that I am cognitively impaired, and that, *all things being equal*, this cognitive impairment would excuse me from blame. Imagine, however, that thirty minutes prior to slugging you I ingested PCP, knowing full well that doing so would likely make me violent. In such a case, my responsibility is not original—I was cognitively impaired when I slugged you. However, this may not get me off the hook. If I am responsible in this case, it is because my responsibility is *derivative*—I am responsible for hitting you because I am responsible for the event that led to my hitting you (ingesting PCP in full knowledge of the likely consequences). We can define the notions of original and derivative responsibility as follows:

derivative responsibility - an agent *S* is derivatively responsible for act

(omission) ϕ at time t_n just in case *S* is responsible for act (omission) ψ at t_{n-m}

⁵⁹ It should be noted that Rosen uses the terms 'moral responsibility' and 'blameworthiness' interchangeably. While this practice is somewhat widespread, I am in the camp that the phrase 'morally responsible actions' denotes a class of actions wider than does the phrase 'blameworthy actions.' What is included in the former class, but not the latter, is the class of actions denoted by the phrase 'praiseworthy actions' (and possibly even attribution-neutral actions).

(where $\phi \neq \psi$ and $n < m$), ψ led to ϕ , and S knew (or reasonably should have known) that ψ would lead to ϕ .⁶⁰

original responsibility – S is originally responsible for act ϕ if and only if S 's responsibility for ϕ is not due to S 's responsibility for any event prior to ϕ (i.e. the presence of *original responsibility* marks the absences of *derivative responsibility*).

With this distinction in hand, Rosen advances the first premise of his master argument:

- (1) If S is responsible for ϕ , then either ϕ is a “locus of original responsibility” or S is derivatively responsible for ϕ , in which case “there exists a locus somewhere in [ϕ 's] causal history.”⁶¹

3.2.2 Ignorance

The next step in Rosen's argument concerns the status of *action from ignorance*. The first thing to note is that there are two types of relevant ignorance: *factual* ignorance and *normative* ignorance. Ignorance of the factual variety occurs when one lacks knowledge of some feature of the world, say the whereabouts of one's car in a parking garage. Ignorance of the normative variety occurs when one lacks knowledge of certain normative principles; these normative principles may be principles of the law, principles of morality, or even principles of prudence. One may be normatively ignorant if one is

⁶⁰ The way in which I have formulated *derivative responsibility* is quite a bit more precise than Rosen does. As Rosen puts it, “in general, X's responsibility for A is derivative when X is responsible for A only because he is independently responsible for B, some prior act or omission.” Putting it this way leaves the connection between A and B too vague.

⁶¹ *Ibid*, p. 299.

unaware that after parking one's car in a large parking garage one should look for nearby landmarks (e.g., signposts) to aid in finding the car later. This is ignorance of an epistemic obligation.⁶²

Sometimes ignorance excuses; sometimes it does not. We can call ignorance that does *not* excuse *culpable ignorance* and ignorance that *does* excuse *excusing ignorance*. If my ignorance of my car's location in the parking garage is a result of my nefarious sister's plot to repark my car on a different level, and the ensuing confusion results in my missing an important meeting, all else being equal I maybe excused for my absence. In such a case, my factual ignorance is *excusing ignorance*. Intuitively, I am not at fault. If, however, my ignorance results from some negligent act—say I rush out of the car after parking it without checking my whereabouts—I may not be excused for missing the meeting, for I have failed to discharge an epistemic obligation. If I am not excused for missing the meeting, this must be because I am culpable for the ignorance (of where my car was parked) that led to my absence.

What distinguishes excusing ignorance from culpable ignorance? This, I believe, is a more difficult question to answer than it may at first appear. One thing to note is that ignorance often results from an act or omission on the agent's part: my ignorance of German is no doubt due to the fact that I have never taken a German class. Of course, ignorance need not so result. I may be ignorant of a certain fact simply because it has never been revealed to me—one might chalk my ignorance of certain esoteric baseball rules to this. There are two points to be gleaned from this. The first is that for any given

⁶² As I understand it, an epistemic obligation is an obligation to put forth reasonable effort in acquiring and maintaining knowledge.

person's ignorance on any given subject there can be many given sources: often our ignorance is overdetermined (my ignorance of certain baseball rules may not only be due to the fact that no one has informed me of them, but also because I have never read the rulebook). This overdetermination is due to the fact that one's ignorance can often be overcome in numerous ways, and thus many factors can be counted as contributing to one's ignorance on any given subject. Take my ignorance of the German language as an example. One reason I don't know German is that I never enrolled in a German language class. Another reason I don't know German is that my mother decided, when I was five, to turn down a job in Germany. Had she decided to accept that job, I likely would have learned German. Yet another reason I don't know German is that the *de facto* language of the United States is not German. If it were, I would likely speak it.

The second point to be gleaned is that almost any given instance of ignorance can be explained counterfactually.⁶³ What this means is that, except in the cases of impossible knowledge, ignorance can be explained by (or traced back to) the occurrence or nonoccurrence of some event. When ignorance is *culpable* this must be because the agent in question is responsible for the event or lack thereof that led to the ignorance. When ignorance is *not culpable*—that is, when ignorance excuses—this is because the agent is *not* responsible for the event (or lack thereof) that led to the ignorance.

We are now in a position to state the second premise:

(2) "Action done from ignorance is never a locus of original responsibility."⁶⁴

⁶³ I say "almost" because certain truths—such as the existence or nonexistence of God, or the number of atoms in the universe—may be impossible to know, and thus ignorance of such truths cannot be explained counterfactually.

⁶⁴ *Ibid*, 300.

It follows from (1) and (2) that

(C1) If S is morally responsible for ϕ and ϕ is an action done from ignorance, then S is *derivatively responsible* for ϕ , in which case there exists a locus of original responsibility in ϕ 's causal history.

In other words, in acting from ignorance a person can only be derivatively responsible, if responsible at all.

3.2.3 *The Structure of Original Responsibility*

We turn now to instances of original responsibility. If the preceding argument is correct, no instance of original responsibility can be an action from ignorance. If this is correct, then what does the structure of original responsibility look like?

In answering this question, Rosen helps himself to the following assumption which we can state as our next premise:

(3) When an act is morally wrong, there is always decisive reason against doing it.⁶⁵

Next we need to help ourselves to the *no blameworthiness without fault* principle:

(4) S can be blameworthy for ϕ -ing *only if* it is morally wrong to ϕ .⁶⁶

I take it, as both Wallace and Rosen seem to, that this principle is intuitively true.⁶⁷ I agree, and shall not attempt to defend it here. From (3) and (4) it follows that

⁶⁵ *Ibid.*, p. 307.

⁶⁶ Cf. Wallace, *Responsibility and the Moral Sentiments*, p. 135; Gideon Rosen, "The Case for Incompatibilism" *Philosophy and Phenomenological Research*, Vol 64, No. 3 (2002), p. 700.

⁶⁷ It should be noted that in "Skepticism about Moral Responsibility" Rosen does not explicitly endorse this principle, though it seems to operate as a background assumption. This principle is explicitly

(C2) If S is blameworthy for ϕ -ing, then S had decisive reason against ϕ -ing.

When it comes to doing what one has most decisive reason *not* to do, there are two possibilities we need to consider: either the person in question *knew* that he or she had decisive reason not to do it, or she did not. If the latter, that is, if, in acting, the agent does not know that she has decisive reason not to do the thing she did, then, by hypothesis, she would be acting under ignorance. However, as we saw above, action done from ignorance cannot be a locus of original responsibility. It follows that for any action which is to count as a locus of original responsibility, it must be the case that the agent, in acting, both knew that doing that thing would be wrong *and* that she had decisive reason not to do that thing. We can restate this as follows:

(5) ϕ can be a locus of original responsibility for S only if (i) S knew that it would be wrong to ϕ , and (ii) S knew that he or she had decisive reason not to ϕ .

Now, under the presumption that doing what one knows one has decisive reason not to do falls under the umbrella of *acting against one's better (best) judgment*, and that *acting against one's better (best) judgment* is to act akratically, the following can be stated as a corollary to 5:

(5c) If ϕ is a locus of original responsibility for S , then in ϕ -ing S must have acted *akratically*.

Because moral responsibility must always bottom out in original responsibility, it follows that “*every culpable bad action must be the causal upshot of a genuinely akratic act or omission.*”⁶⁸ Or, to put it another way:

(C3) *S* can be blameworthy for ϕ only if *akrasia* occurs somewhere in ϕ 's etiology (i.e. causal history).

It should be noted that (C3) is the most important lemma in Rosen's argument. It is this that leads to his soft skeptical upshot, and it is the element of Rosen's argument that I will exploit in the argument of section 4. It should also be noted that by itself, there is no *prima facie* reason to think that (C3) will lead to a skeptical conclusion in regards to blameworthiness. It is this fact, I think, that makes Rosen's soft skepticism so potentially powerful.

3.2.4 The Soft Skeptical Upshot

Rosen admits that by itself (C3) does not warrant skepticism about moral responsibility. The skeptical upshot comes into play when we try to assess whether any given action is blameworthy. Given (C3), in order to ascertain whether *S* is blameworthy for ϕ -ing, we must ascertain whether ϕ itself was an instance of *akrasia*, or whether it was the causal upshot of an akratic act.⁶⁹ In other words, if we are to judge that *S* is in fact blameworthy for ϕ -ing we must be able to confidently judge that either ϕ itself is an instance of *akrasia* or that it was the casual upshot of an akratic act.

⁶⁸ Rosen, “Skepticism about Moral Responsibility,” p. 307.

⁶⁹ This, of course, is only a necessary condition. That an act is akratic does not, by itself, make it blameworthy (though it may be sufficient to make it rationally *criticizable*).

Rosen does not think that making such judgments is impossible *per se*. Rather, he thinks that, “given the opacity of mind—of other minds and even one’s own—it is almost always unreasonable to place significant confidence in such judgments.”⁷⁰ There are two main factors that make identifying genuine blameworthiness problematic. The first is that the relevant facts are often inaccessible.⁷¹ Consider an action done from ignorance: a student forgets to turn in an assignment. If she is blameworthy for so doing, the student must be culpable for the ignorance under which she acts. But if she is culpable for the ignorance under which she acts, the ignorance must be the causal upshot of some “prior culpable act or omission.”⁷² It might be the case that we know the student in question rather well. Imagine that she has a history of absentmindedness, especially when it comes to turning school assignments in on time, and that we have personally told the student to keep a calendar of due dates and check it every Monday, Wednesday, and Friday. Let’s pretend for simplicity sake that checking the calendar would have been sufficient for her to turn her assignment in on time. From the fact that she failed to turn her assignment in, we can infer that she failed to check the calendar. Now, if she is blameworthy for forgetting to turn her assignment in—an action (omission) done from ignorance—and this was a result of her failure to check her calendar, she must be culpable for this failure. Why did the student fail to check her calendar? She either forgot to check the calendar (action from ignorance) or she didn’t. If the former, then we

⁷⁰ Rosen, “Skepticism about Moral Responsibility,” 308.

⁷¹ *Ibid.*

⁷² *Ibid.*

need to ascertain if her ignorance is culpable; if the latter, then we have to ascertain why she didn't check her calendar, though she knew she should have.

We can stop the exercise here; the point, I trust, has been made: when ascertaining culpability we will often have to search through a long etiology. In practice, this will rarely be possible. This, however, is just the start of our problems. The other major problem concerns distinguishing *genuine akrasia* from a common imposter: weakness of will. It was suggested in section 3.1 that we can distinguish two distinct phenomena that often go under the moniker 'weakness of will.' The first can be characterized as a judgment-shift, or the kind of phenomena that occurs when one gives in to temptation, as when one gives in to the temptation to eat a piece of chocolate cake for desert. What often occurs in such a process is that one shifts from a judgment such as "*I really should not eat cake tonight*" to "*Eating one piece of cake tonight won't hurt.*" The individual then proceeds to eat a piece of cake. This sort of phenomena is distinct from what is usually identified as *akrasia* (but sometimes goes by 'weakness of will'), acting against one's best judgment. What distinguishes cases of *akratic* action from cases of weakness of will is that with the former no judgment-shift occurs. That is, an *akratic* actor may form the judgment "*I really should not eat cake tonight*" proceed to eat the cake, but not shift to a judgment such as "*Eating one piece of cake tonight won't hurt*" prior to the act. The problem is this: weakness of will, understood as judgment shift, is technically acting under ignorance, and so if it is to be culpable, we must trace it back to a culpable act, namely an *akratic act*. But in practice it is next to impossible to discern weakness of will from *akratic* action. Due to what Rosen calls "the opacity of

mind,” this is not only exceedingly difficult in the case of the first-person, but even more so when it comes to the third-person case. So says Rosen: “It is almost always unreasonable to place significant confidence in such a judgment.”⁷³

Rosen’s take-home point is not that agents are *in fact* never morally responsible, or that one can never truthfully make an ascription of moral responsibility, but rather that because of the difficulty involved in accurately identifying morally responsible action, it is unlikely that one can ever *confidently* make such an ascription.

3.3 FitzPatrick’s Response

In the seven years since its publication, with the exception of two articles published in *Ethics*, Rosen’s article has received surprisingly little attention. The first of these two articles, by William FitzPatrick, offers a counter-argument to Rosen.⁷⁴ The second of these articles, by Neil Levy, defends Rosen against FitzPatrick.⁷⁵ In this section I will relate FitzPatrick’s rejection of Rosen. In the following section I will relate Levy’s response to FitzPatrick.

⁷³ *Ibid*, p. 308. To compound the problem, Rosen notes that there are intermediary states in between *akrasia* and weakness of will: “cases in which the agent *suspends* his original judgment without quite rejecting it, or cases in which the it is simply indeterminate as the agent acts whether he in fact believes that all things considered he should do [what he does].” *Ibid*, p.309.

⁷⁴ “Moral Responsibility and Normative Ignorance: Answering a New Skeptical Challenge” *Ethics* 118, No. 4 (July 2008): 589-613.

⁷⁵ Levy, Neil, “Culpable Ignorance and Moral Responsibility: A Reply to FitzPatrick,” *Ethics* 119, No. 4 (2009): 729-741.

3.3.1 *The Opacity of Mind?*

FitzPatrick offers two arguments in response to Rosen. The first concerns what I have called *the soft skeptical upshot*, that is, the claim that one cannot confidently ascribe responsibility to any particular individual due to the fact that genuine *akrasia* is hard to identify. FitzPatrick finds Rosen's claims here to be dubious and easily refutable. Despite this, FitzPatrick still thinks that the power of Rosen's argument is due to its defense of (C3).⁷⁶ If this part of the argument is sound, reasons FitzPatrick, then we would find that a large category of *prima facie* morally responsible action, in particular cases involving normative ignorance, would be ruled out as morally responsible insofar as they lack a locus of *akrasia*.

While, like FitzPatrick, I am more interested in the (C3) lemma than the soft skeptical upshot, I will start with FitzPatrick's response to the latter before moving on, in 3.3.2, to his response to the lemma.

Rosen's soft skeptical upshot arose out of supposed epistemic difficulties associated with identifying genuine cases of *akrasia*. FitzPatrick attempts to undermine the soft skeptical upshot by arguing that instances of genuine *akrasia* are not nearly as difficult to identify as Rosen claims. His strategy in doing so is to tell two stories. The first story offers an account of how we can identify genuine instances of *akrasia* in the first person, while the second story offers an account of how we can identify such instances in the third person.

⁷⁶ As a reminder, (C3) states, "*S* can be blameworthy for ϕ only if *akrasia* occurs somewhere in ϕ 's etiology (i.e. its causal history)."

As we saw, the primary difficulty that Rosen cites in attributing *akrasia* to agents in *prima facie* blameworthy action comes in the form of distinguishing cases of “ordinary weakness of will” from cases of *akrasia* proper.⁷⁷ Rosen claims that, even in the case of the first person, it is unlikely that we can make anything like a confident identification. FitzPatrick presents evidence that Rosen’s appeal to epistemic difficulty is overblown.

FitzPatrick’s first piece of evidence is a “quasi-autobiographical reflection” concerning the phenomenology of what we might call *going wrong*.⁷⁸ I will quote FitzPatrick’s autobiographical reflection in full:

I often know such things as that I really shouldn’t be digging into a heaping bowl of full fat Belgian Chocolate ice cream given my cholesterol levels and the fact that I’ve had two helpings already this week. And I know this perfectly well even at the moment I am doing it, as it is transparently imprudent according to standards I myself accept, even taking all things—such as my present enjoyment—into account. I’m just not sufficiently motivated by these normative thoughts, instead giving into gustatory temptation. Each instance may be no big deal in itself, given the long term and statistical nature of the health issues, but I nonetheless know that it is not what I should be doing here and now.⁷⁹

⁷⁷ FitzPatrick states in passing that the so-called cases of “ordinary weakness of will” to which Rosen refers might be better called “weakness of conviction.” I side with Rosen on this terminological dispute. Such cases of judgment shift in the face of temptation serve as an appropriate antipode to strength of will or will power, that is, standing firm in the face of temptation.

⁷⁸ By “going wrong” I mean simply doing what one should, all things being equal, not do. I do not mean, as shall become clear, necessarily going *morally* wrong.

⁷⁹ FitzPatrick, “Moral Responsibility and Normative Ignorance,” p. 594-595.

It is not clear to me how effective FitzPatrick's pseudo-autobiographical reflection is as a counterexample. First, that I recognize good reason not to do a certain thing does not mean that doing that thing in full recognition of such reason constitutes *akrasia*. The fact that I have high cholesterol is presumably just a *pro tanto* but not *perfect* reason to refrain from eating ice cream.⁸⁰ FitzPatrick admits that, when indulging in ice cream, he *does* recognize considerations in favor of doing so (viz., the "enjoyment" that it would bring). In order for his eating ice cream to count as an *akratic* act, it must be the case that he takes his considerations in favor of *refraining* from ice cream eating to outweigh, all things considered, the reasons that weigh in favor of *eating* ice cream. Unfortunately, FitzPatrick's "pseudo-autobiographical reflection" is too under-described to discern anything one way or the other. FitzPatrick admits of two competing considerations occurring in his mind at the time, that concerning his cholesterol, and that concerning his assured enjoyment of the ice cream. However, it is not clear if, at the time of action, one *really* outweighs the other. While it *is* clear that at the time of action FitzPatrick *does* think eating ice cream would be imprudent, whether he *actually* takes this to outweigh the considerations in favor of eating it— such as the sure bliss that eating chocolate ice cream brings—is uncertain. As hard as it is to tell how the reasons weigh against each other in the case of FitzPatrick's "pseudo-autobiographical reflection," we can expect it to be even harder in the case of real-life reflections. FitzPatrick admits of the possibility of self-deception. However, he insists that:

⁸⁰ For the distinction between *pro tanto* and *perfect* reasons see John Broome, "Reasons," in *Reasons and Value, Themes from the Moral Philosophy of Joseph Raz*, edited by R. Jay Wallace, et al (Oxford: Oxford University Press, 2004): 28-55.

the fact that such cases of self deception are possible does not imply that we cannot often know that this is not what is happening and that we are instead just acting badly out of self-indulgence, fear, laziness, greed, resentment, and so on. Nor is it difficult to find good evidence for this.⁸¹

It should be noted that recognizing that one has acted out of self-indulgence, fear, laziness, greed, or resentment is not sufficient to determine that one is thus acting *akratically*. Often, when looking back on actions resulting, for example, from fear, we will self-describe our fear as not sufficiently justifying our actions and take ourselves to have acted against our better judgment. In addition, as bad as self-indulgence, fear, laziness, greed, and so on are as far as reasons for acting go, one can act according to such reasons yet still be doing exactly that which one had judged one has most reason to do. *Enkratic* action⁸² need not be fully rational action.

But what of the other evidence FitzPatrick provides? FitzPatrick claims that when acting *akratically*, we will often “experience guilt or shame.”⁸³ The suggestion here is that such feelings often arise after acting *akratically*. When we do what we take ourselves to have reason to do, says FitzPatrick—even when acting under ignorance—we generally don’t feel guilty or shameful about it.

While FitzPatrick admits that such evidence is “fallible,” he insists that “the presence of guilt of shame at the time of acting is ... often good evidence of *akrasia*.”⁸⁴ I find the connection to be spurious. The nature of guilt and shame is a bit outside the

⁸¹ FitzPatrick, “Moral Responsibility and Normative Ignorance,” p. 595.

⁸² Acting according to one’s best judgment.

⁸³ *Ibid.*, p. 595.

⁸⁴ *Ibid.*

scope of this work, but it seems likely that the extent to which one experiences these emotions will have more to do with certain features of one's biography (e.g. one's childhood) than the motivational structure of any given action. Guilt and shame are often experienced when one believes one has done (or is doing) something wrong. This belief can likely be either conscious or subconscious, and can certainly be connected to *enkratic* actions. I might, for example, feel guilty or shameful for having premarital sex, but this might very well be because I made a pledge of monogamy. My guilt and shame are more likely tied to the fact that I broke a promise than to any akratic act on my part. Any connection there may be between guilt and shame on the one hand and *akrasia* on the other is certainly too flimsy to base any kind of confident judgment.

FitzPatrick's claim for the possibility of confident third-personal attributions of *akrasia* is really only as strong as his claim for the possibility of confident first-personal attributions of *akrasia*, for his defense of such confident attributions largely depends on the self-reporting of *akratic* episodes by others. Says FitzPatrick: "Sometimes people are honest about their feelings and will come out and admit that they knew better, even at the time of acting, and felt ashamed even as they acted."⁸⁵ Why might this be? Here is FitzPatrick's answer:

[A]lthough it might be tempting to plead normative ignorance [i.e. that one didn't know that what one did was wrong], it would actually be more embarrassing to do so and often simply incredible. Someone who has committed a particularly egregious act of adultery, for example, and is trying to explain himself to his

⁸⁵ *Ibid*, p. 596.

spouse might find it hardly more credible to plead normative ignorance than it would to plead circumstantial ignorance (“I didn’t realize it was wasn’t you!”),⁸⁶ and in any case it might reflect so badly on his values if he really thought at the time that the act was justified that it wouldn’t help. So people are sometimes honest with themselves about having acted *akratically*...and are honest with others about it too.⁸⁷

Rather than find reason to think we can form confident third-personal attributions of *akrasia*, FitzPatrick’s account lends support to Rosen’s skeptical position. FitzPatrick claims that it would often be more “embarrassing” to plead ignorance than *akrasia*. While this is certainly evidence that one will be *inclined* to “admit” *akrasia*, it is certainly not evidence that one who has so “admitted” has *in fact* acted in such a way. If it would be more embarrassing or less efficient to plead ignorance (even if the truth) than “admit” *akrasia*, then this would be reason to “admit” *akrasia* even when one has not so acted. It is even reason for one to *believe* that one has acted *akratically* when one has in fact acted out of ignorance. In other words, what FitzPatrick has given us is just more evidence for skepticism regarding the possibility of forming confident attributions of *akrasia*.

⁸⁶ The more likely (true) plea of ignorance at play in such cases would be “I didn’t think I would get caught!” When such considerations are involved in one’s deliberation, it is unlikely that one acted *akratically*, even when one judges something to be wrong. One is likely taking the fact that one won’t get caught, together with other considerations (such as, for example, the pleasure it will bring), as sufficient reason to do a certain thing. In practice, however, simply “admitting” *akrasia* rather than pleading this sort of ignorance will be a better strategy, but not reason why we should think in such cases that one has actually acted *akratically*.

⁸⁷ *Ibid*, p. 596.

The problem we run into when it comes to identifying instances of genuine *akrasia* is more than just the opacity of mind that Rosen eludes to. In addition to this opacity—whatever it amounts to—we find the presence of distorting emotions (shame and guilt) tied to certain actions as well as social pressures to admit that such actions were akratic in nature. Combined with the fact that the *akrasia* we are looking for is often buried deep in an act’s etiology, FitzPatrick’s considerations only prove to support Rosen’s soft skeptical upshot, rather than undermine it.

3.3.2 *Rejecting the Lemma*

Despite my criticism of FitzPatrick’s attempt to undermine Rosen’s soft skeptical upshot, I am ultimately not concerned with the upshot’s success. The hard skeptical upshot that I defend in section 4 is not skepticism of the epistemic variety that Rosen defends. It does not depend on any difficulty associated with identifying genuine instances of *akrasia*. Rather, the hard skeptical upshot I defend argues that *blameworthiness* is impossible. The important premise of my argument has it that according to the norms internal to the practice of blame—the norms articulated by Wallace in *Responsibility and the Moral Sentiments*—*akrasia* satisfies an *exempting condition* from blame. However, in order for this claim to produce the hard skeptical upshot I desire, Rosen’s important lemma (C3) must be true. For convenience, I’ll repeat that lemma here:

(C3) *S* can be blameworthy for ϕ only if *akrasia* occurs somewhere in ϕ ’s etiology (i.e. causal history).

One can see how (C3), together with my claim that *akrasia* excuses, will produce a hard skeptical upshot by considering the contrapositive of (C3):

(C3cp) If *akrasia* does not occur in ϕ 's etiology, then S cannot be blameworthy for ϕ .

In order for my hard skeptical upshot to go through, I must defend (C3) against possible attacks. The rest of this subsection shall present an objection to (C3) leveled by FitzPatrick after attempting to dispense of the soft skeptical upshot.

While FitzPatrick believes that the objections he raises to Rosen's soft skepticism are sufficient to undermine it, he recognizes that the truth of (C3) *itself* would have startling consequences for normal moral judgments. The most important such consequence concerns the culpability of ignorance. We saw above that, according to Rosen, action from ignorance is culpable only if one is culpable for the ignorance out of which one acts. And one is culpable for the ignorance out of which one acts only if one is culpable for the act or omission that gave rise to the ignorance. In order for ignorance to be culpable, it must bottom out in *akrasia*. When considering cases of normative ignorance, that is cases where one is acting badly but does not believe that one is so acting, this line of reasoning will often conflict with common intuitions that such actions *are*, in fact, culpable. Consider the following case, developed by FitzPatrick:

Ruthless Businessman: "Mr. Potter, a powerful businessman...holds false moral views. He takes certain business practices—such as liquidating Bailey's Building and Loan and sticking it to the poor families of Bedford Falls—to be 'permissibly aggressive,' when in fact they're 'reprehensibly ruthless.' This leads

him to do bad things, though he doesn't understand that he's acting badly, which means that he's acting out of a certain kind of ignorance. He's fully aware of all the circumstances, but he applies flawed normative principles or weightings and comes up with bad decisions."⁸⁸

When Mr. Potter liquidates Bailey's Building and Loan, he does so under the guise that it is "permissibly aggressive." In doing so, Mr. Potter is not acting *akratically*. Rather, he is acting under normative ignorance. For FitzPatrick, the question of whether Potter (or anyone for that matter) is culpable for his ignorance depends on how we answer the following question:

(R) "What, if anything, could the agent reasonably (and hence fairly) have been expected to have done in the past to avoid or to remedy that ignorance?"⁸⁹

When it comes to answering this question, we can understand Rosen as maintaining that Potter could have reasonably been expected to avoid or remedy his ignorance (and thus be culpable for possessing it) only if he was in possession of the relevant facts and norms pertaining to the correct or appropriate way to manage or maintain his knowledge, such as through the appropriate amount of self-reflection or gathering of information. If Potter was not, at some time, in possession of these relevant facts or norms, then, according to Rosen, it is not reasonable to have expected Potter to have done anything to remedy his ignorance. Presumably, in order to do something to remedy his ignorance, Potter would have had to of either *stumbled* into remedying it, or else been aware of the

⁸⁸ FitzPatrick, "Moral Responsibility and Normative Ignorance," pp. 599-600. This, of course, is based on Frank Capra's film, *It's a Wonderful Life* (1946).

⁸⁹ *Ibid.*, p.603.

relevant facts or norms requiring him to remedy his condition. Since it is too much to *expect* one to simply stumble into something, Potter can only reasonably be expected to have remedied the state of his knowledge if he was aware that should, but failed to discharge this duty.

FitzPatrick thinks that Rosen's *akrasia* condition on normative ignorance is too strong. It is likely that no such akratic state occurred in the etiology of Potter's ignorance. It may have been the case that, for example, he was raised to be insensitive to the suffering of others and to simply dismiss competing viewpoints as "socialist or liberal without honest reflection open to the possibility that they may point to blind spots in his own views."⁹⁰ Despite this, it is likely, given that Potter was raised in the same community of Bedford Falls as George Bailey, the morally upright pillar of the community, that "the opportunity for improved normative understanding was clearly present in his social context."⁹¹ That is, Potter had the same opportunity for and access to a normative education as the other residents of Bedford Falls, but Potter dismissed such moral growth as a sign of weakness or stupidity. Success, in Potter's mind, requires a cold heart and unpopular decisions. While his neighbors in Bedford Falls would likely be more popular and liked, he would be more successful and possibly control the town some day.

If such was the etiology of the ignorance that lead to Potter's immoral business practices, we can say that it is likely it never occurred to Potter that he *should* seriously consider alternative viewpoints and not simply dismiss them in a cavalier manner.

⁹⁰ *Ibid*, p. 604.

⁹¹ *Ibid*, p. 603-604.

Maybe Potter knew that what he was doing was “morally wrong,” but it is likely that Potter thought morality to be nothing other than a social construct, or a system of slave morality—maybe he misread Nietzsche in high school and viewed himself as following the lead of the *Übermenschen*—and that a moral education was not any more important, or different, than an education in etiquette. The point is that many of Potter’s business practices, while resulting out of normative ignorance, do not bottom out in *akrasia*.

Does this mean that the ignorance under which Potter acts is not culpable, and thus that he is not blameworthy for his immoral business practices? FitzPatrick thinks that the answer to this is obviously and intuitively “no.” He provides the following to support this judgment:

- (A) “There were no relevant limitations in his social context or in his capabilities that should have made the necessary broader reflection and information gathering impossible or unreasonably difficult for him.”⁹²
- (B) “The failure of adequate reflection and information gathering was instead the result of voluntary exercises of vice such as overconfidence, arrogance, dismissiveness, laziness, dogmaticism, incuriosity, self-indulgence, contempt, and so on.”⁹³
- (C) “He could thus reasonably have been expected to take steps that would have eliminated that ignorance, by refraining from exercising those vices and instead taking advantage of the epistemically relevant opportunities available to him.”⁹⁴

⁹² *Ibid.*, p. 605.

⁹³ *Ibid.*

⁹⁴ *Ibid.*

As far as I understand FitzPatrick, (A) and (B) are meant to be answers to the question (R) quoted above. What is not clear is whether (A) or (B) alone are meant to be sufficient for one's ignorance to be culpable, or whether neither is alone sufficient, but both together *would* be sufficient. It should be noted that FitzPatrick does not, of course, think (A) and (B) are *necessary* for ignorance to be culpable, for he does not dispute that *akrasia* can be a source of culpability. Whatever the case may be, it is not ultimately important what the status of (A) and (B) are, for if he is correct that (A) and (B) are sufficient for Potter's ignorance to be culpable, and *akrasia* does not ground either (A) or (B), then he will have successfully defeated (C3).

FitzPatrick notes that Rosen will likely respond to this by pointing out that we can only be culpable for those actions resulting from our vices if we are in turn culpable for possessing or excising said vices. In light of this, Rosen will draw the following dilemma: either Potter was aware that he was not acting well in his vice-related action—in which case he would be acting akratically—or else he thought that he *was* acting well—in which case he would be acting from ignorance. If the former, then we do not, in fact, have a counter-example to Rosen's lemma. If the latter, then, according to Rosen, we must yet again assess whether Potter's ignorance is culpable, which will require investigating whether or not Potter failed to carry out his epistemic duties, and if so, whether this failure is *itself* culpable.

While FitzPatrick agrees that in order for Potter to be culpable for his bad business practices, he must be culpable for excising or possessing the vices that led to his aggressive acts, FitzPatrick wants to resist the move “of passing the buck of

responsibility every time we come across a bad choice involving ignorance.”⁹⁵ Even if Potter exercised his vices out of ignorance, and this ignorance does not bottom out in *akrasia*, so long as Potter’s vice-related actions occurred “in a context where he could reasonably have been expected to know better and to do a better job of informing himself morally,” this is enough, according to FitzPatrick, for Potter to be blameworthy for his actions.⁹⁶

At this point I’d like to pause to point out a couple of things that have arisen in the exchange between Rosen and FitzPatrick. The first is that both Rosen and FitzPatrick seem to agree that blameworthiness concerns holding others to *reasonable* expectations. When those expectations are violated, the agent is liable to blame—this squares nicely with the account of blameworthiness that we find from R. Jay Wallace.⁹⁷ The second point concerns the nature of their disagreement. For Rosen, it is reasonable to hold a person to an expectation *p* only if that person, in acting, was *aware* that she *ought* to *p*, or if she was not aware that she ought to *p*, then she violated some prior obligation that she *was* aware of having and that violating *this* obligation led to the ignorance in question. We can call this condition the *strong condition on reasonable expectation*. For FitzPatrick, it is reasonable to hold a person to an expectation *p* so long as (a) she has the capacity to *p*; (b) she has the opportunity to *p* given her social context; and (c) her failure to *p* resulted either from exercising a vice or from an *akratic* act. We can call this the *weak condition on reasonable expectation*.

⁹⁵ *Ibid.* p 606.

⁹⁶ *Ibid.*

⁹⁷ Cf. R. Jay Wallace, “Responsibility and the Moral Sentiments,” and section 2.3 above.

We can understand the difference between the *strong condition* and the *weak condition* in terms of requirements on knowledge. According to the *strong condition* we can reasonably hold a person to an expectation only if he or she fails to satisfy an obligation that he or she was aware of having.⁹⁸ According to the weak condition, however, we can reasonably hold a person to an expectation so long as she employed this kind of knowledge elsewhere in her life.⁹⁹

It will be helpful in unpacking the difference between the strong and weak condition by turning back to **The Case of the Ruthless Businessman**. Let's stipulate that in liquidating Bailey's Building and Loan, Mr. Potter, the antagonist of our story, did so because he correctly believed that it would improve his bottom line, and improving his bottom line is viewed as the most important value in his life (he may even value his bottom line over his health). Presumably, given the importance Potter places on improving the bottom line, he understands the importance of hearing opposing points of view when it comes to making important business decisions. If this is so, says FitzPatrick, it is not unreasonable to expect Potter to put this practice to work (viz., that of hearing out opposing viewpoints) in his moral life. We might imagine that were he to

⁹⁸ This raises an interesting question about derivative responsibility. Imagine that I have just committed a moral transgression ϕ , but that I ϕ -ed out of ignorance. According to Rosen, I am responsible for ϕ only if I am responsible for the ignorance that results in ϕ . Call this state of ignorance s . I am responsible for s only if I failed to fulfill some epistemic obligation, call it o , that I was aware I had. Imagine that I did fail to fulfill o and knew that I had an obligation to o , but that I was wholly unaware that failing to o would or could lead to ϕ . While it may seem reasonable to hold me responsible for failing to o , it does not seem as reasonable to hold me accountable for ϕ on account of my failing to o , for while I was aware that I had an obligation to o , I was not aware that failing to o might lead to my ϕ -ing. This line of reasoning, if correct, would lead to the following principle: S can be derivatively responsible for ϕ only if S violated an *epistemic* obligation, o , in full awareness that she had an obligation and knew that in failing to o she would ϕ or would likely ϕ and that ϕ is something that she ought not do. This condition seems to be much stronger than that defended by Rosen.

⁹⁹ This, I think, is what FitzPatrick has in mind when he talks about "capacities".

do so, he might come to realize that liquidating Bailey's Building and Loan is simply immoral, and that this consideration overrides his bottom line. Importantly, it was perfectly within Potter's capability to do so, given the fact that he employs this method of information gathering in another sphere of his life. But even though Potter did not believe that he had to extend this practice to the moral domain, he did have the capacity and the opportunity to do so (given his social context) and, we can imagine, his obsession with wealth (clearly a vice!) was somehow involved. Because of this, it is reasonable to expect Potter to extend the practice of hearing out opposing sides into the moral domain. And because he failed to do so despite his capacity and opportunity to do so, he is culpable for any ignorance to which it leads and any actions that results from this ignorance.

If FitzPatrick is right, if the weak condition of reasonable expectation is all that is required for ignorance to be culpable, then Potter is blameworthy for liquidating Bailey's Building and Loan despite the fact that *akrasia* occurs nowhere in the etiology of that act. And if this is correct, then Rosen's lemma is false.

3.4 Defending Rosen's Lemma

Because the skeptical argument I supply in section 4 exploits Rosen's lemma, I must defend against FitzPatrick's objection. The argument advanced here is adapted from that supplied by Neil Levy.¹⁰⁰ It relies on the intuitive assumption that one cannot reasonably expect another to act irrationally. It will be argued that FitzPatrick's weak

¹⁰⁰ Neil Levy, "Culpable Ignorance and Moral Responsibility," *Ethics* 119, No. 4 (2009):729-741.

condition on reasonable expectation amounts to a denial of this claim, and thus should be rejected. What follows is a brief summary of this argument (3.4.1), followed by a possible objection (3.4.2), and then my reply to that objection (3.4.3).

3.4.1 Levy's Argument

It will be helpful to begin by restating the lemma in question:

(C3) *S* can be blameworthy for ϕ only if *akrasia* occurs somewhere in ϕ 's etiology (i.e. causal history).

FitzPatrick rejects this claim. He defends the following condition on blameworthiness:

(F) *S* can be blameworthy for ϕ only if in ϕ 's etiology there were some means such that, had *S* taken those means, *S* would likely have not ϕ -ed, *and it was reasonable to expect S to have taken those means.*¹⁰¹

The key point of disagreement between FitzPatrick and Rosen is with respect to what can be reasonably expected of an agent. For Rosen, *S* can be reasonably expected to ϕ if and only if *S* is aware that she ought to ϕ . For FitzPatrick, *S* can be reasonably expected to ϕ if and only if *S* has the capacity and opportunity (given her social context) to ϕ .¹⁰²

Above, we called these two conditions, respectively, the *strong condition* and *weak condition* on reasonable expectation.

To defeat FitzPatrick, one needs to argue that the weak condition can be satisfied (i.e. that someone can have the opportunity and the capacity to do a certain thing in the

¹⁰¹ I construe this as a necessary and not sufficient condition on blameworthiness because there is certainly at least one other condition required for blameworthiness: the act must be a moral transgression.

¹⁰² Here ϕ is not necessarily an act of original responsibility, it may be an epistemic obligation, or even a state of knowledge.

way that FitzPatrick understands this) without there being a reasonable expectation for action. And without a reasonable expectation, failing to so act cannot result in blameworthiness later in the causal history.¹⁰³

One way to argue that there can be no reasonable expectation for a certain course of action is to show that it would be irrational to take the course of action in question. This is the route that Levy takes in his argument. Before diving into this, however, let's revisit, from section 2.3.2., the different senses of the notion of a "reasonable expectation." There are two ways to understand this phrase, one epistemic, the other normative. One might say that there is a reasonable expectation that the winner of this year's mayoral race in San Francisco will be a member of the Democratic Party. This kind of reasonable expectation is of the epistemic variety. This expectation, which concerns what one is justified in *believing* (or *expecting* to happen), is not the kind of expectation that concerns Rosen and FitzPatrick. The kind of reasonable expectation that concerns them is the reasonable expectation of the *normative* variety. This is the expectation of *obligation*.

Because the disagreement between Rosen and FitzPatrick centers around a disagreement about how to understand reasonable expectation, simply stipulating a definition of what counts as reasonable expectation will be question begging and simply won't do. The route that I shall pursue—a route similar to that taken by Levy—will be to put forth a notion of irrational action. With this in hand I will then argue that it would

¹⁰³ We can call this principle the "no blameworthiness without expectation" principle. This principle is similar to the "no blameworthiness without fault" principle. The difference is that, in the former the expectation broken need not be a moral expectation, whereas in the latter, the fault *is* understood to be a moral transgression.

be irrational for Potter to do what is required of him in order to satisfy FitzPatrick's weak condition on reasonable expectation.

Now, we can distinguish between an *objective* and a *subjective* notion of rationality.¹⁰⁴ Often, when we talk about what it is rational for a person to do, we are talking about what it is *objectively* rational for a person to do. Here, we are concerned with what a person, all things considered, has most reason to do. When speaking of what it is rational to do in this sense of the notion, we might say something like "Well, the rational thing for *S* to do in such a situation is *x*."

Often, however, agents do not have access to all their reasons for acting.¹⁰⁵ That is to say, it will often be the case that there are reasons that favor a certain course of action for a certain agent, but that the agent in question is not *aware* of those reasons.¹⁰⁶ It might, for example, be the case that the fact that it is raining outside is a reason in favor of my bringing an umbrella to campus today. In light of this fact, one might say

¹⁰⁴ Levy marks the distinction this distinction as between *external* and *internal* notions of rationality. The only difference here is a terminological one. However, using the terminology that I do will keep confusion at bay when we start discussing internal and external reasons.

¹⁰⁵ Several things need to be said here. First I shall be using the term 'reason' in its *normative* rather than *explanatory* sense. That is, *R* is a reason for *S* to ϕ just in case ϕ counts as a consideration in favor of ϕ -ing (Cf. T.M. Scanlon, *What We Owe to Each Other* [Cambridge: Harvard University Press, 1998], especially chapter 1). Second, that *S* has a reason to ϕ does not imply that *S* "ought to ϕ ". Rather, I will be using 'reason' in the '*pro tanto*' rather than '*perfect*' sense (cf. John Broome, "Reasons" in *Reason and Value, Themes from the Moral Philosophy of Joseph Raz*, edited by R. Jay Wallace, et al, [Oxford: Oxford University press, 1994]): *R* is a *pro tanto* reason for *S* to ϕ if and only if *R* weighs in favor of *S* ϕ -ing. Given this understanding of 'reason,' it is perfectly consistent for *S* to have a reason to ϕ and for *S* to have a reason to refrain from ϕ -ing. Third, my argument is neutral with respect to the internal/external reasons dispute. While my discussion will obviously be consistent with reasons understood in the external sense, I believe it is also consistent with reasons in the internal sense. (*R* is a reason for *S* to ϕ in the internal sense if and only if *R* is in *S*'s motivational set or can be derived from *S*'s motivational set; for *R* to be a reason for *S* to ϕ in the *external* sense, *R* need not be in *S*'s motivational set.) (Cf. Bernard Williams, "Internal and External Reasons" in *Moral Luck* [Cambridge, UK: Cambridge University Press, 1981]: 101-113; and Nomy Arpaly, "On Acting Rationally Against One's Best Judgment" *Ethics* 110, No. 3 [2000]: 488-513.)

¹⁰⁶ In other words, 'access,' as I understand it, is an epistemic relation between an agent and her reasons: *S* has access to a certain reason *R* to ϕ if and only if *S* is aware that *R* counts as a consideration in favor of ϕ -ing.

“the rational thing for you to do is to bring an umbrella today.” Imagine, however, that I am not aware of the fact that it is raining (I may have my blinds shut), and believe, instead, that it is sunny outside. As such I decide not to bring my umbrella. In such circumstances, I am doing precisely that which I believe I have most reason to do. As it happens, what I believe is false, but were my beliefs true, I *would* be doing what I would have most reason to do. While my decision to bring my umbrella may not be in line with what I *in fact* have most reason to do, and thus not objectively rational, it is in line with what I *take* myself to have most reason to do. In such circumstance we may say that I am acting *subjectively* rational.¹⁰⁷

In a certain respect, we ought to always do what we have most reason to do. Some might even say that this is a trivial truth. In a another respect, however, sometimes it would simply be *irrational* to do what we have most reason to do. Imagine I *did* bring my umbrella today, despite the fact that I thought it was sunny. Unaware of the internal state of my mind, you might ask me, “How’d you know to bring your umbrella?” To this, were I to be truthful, I would reply: “Well, I thought it was sunny, but I brought my umbrella anyway.” It would only be natural for you to be dumbfounded, but quite *unnatural* for you to respond, “Well that was quite rational of you; after all, given that it’s raining today, you had most reason to bring your umbrella.”

As this example illustrates, objective and subjective rationality come apart. It does not always make sense for a person to do what she *in fact* has most reason to do.

¹⁰⁷ It may be that in order for one to act subjectively rational one’s beliefs must also be theoretically rational. This seems right to me, but for the purposes of this paper I wish to remain neutral with respect to this. Cf. Ralph Wedgewood, “Choosing Rationally and Choosing Correctly,” in *Weakness of Will and Practical Rationality*, edited by Sarah Stroud and Christine Tappolet (Oxford: Oxford University Press, 2003): 201-229.

What it is objectively rational for a person to do will sometimes be an action that is subjectively irrational to perform.

So the question we've come to is this: can we reasonably expect of a person that she do something which it would be subjectively irrational for her to do? One can, of course, try to show a person that she is mistaken in what she thinks she has most reason to do. But in doing so, one is not really expecting of the person that she act subjectively irrational, so much as one is trying to show the individual that she is wrong in her beliefs. The question is not whether we can get people to change their mind about what they have most reason to do, but a question about what we can expect of people given the reasons of which they are aware. I believe the answer to the question must be negative. It is unreasonable to expect people to do something which it would be subjectively irrational to do. Consider the following story:

Bank Robbery: Jill is a teller at bank. Every day Jill wakes up at 7 a.m. and, prior to driving to work, she watches the local morning news as she drinks her coffee and eats her breakfast. Today, however, Jill's T.V. stopped working. As a result she can't watch the morning news as she drinks her coffee and eats her breakfast. She leaves her house at the usual time and drives to work. However, had she watched the morning news, she would have seen that her bank was being robbed.

Presumably, given that her bank is being robbed, Jill has most reason not to drive to work today. However, Jill is unaware of the robbery in progress and thinks that today is business as usual.

Is it reasonable to *expect* Jill to stay home today? We can imagine two kinds of scenarios that result in Jill staying home. The first involves Jill conceiving of a reason sufficient for her to stay at home. The second involves Jill *akratically* staying at home, that is, staying at home even though she believes she has most reason to go to work.¹⁰⁸ Certainly we can't *expect* Jill to act akratically. Thus, if it is reasonable to expect Jill to stay home, it must be that we expect her to conceive of a reason to stay home, and that it is reasonable to expect her to conceive of such a reason.

I don't think that, given the circumstances Jill has found herself in, many will find it reasonable to expect her to stay home. I believe, however, that Jill's situation is similar in structure to Mr. Potter's. If it is not reasonable to expect Jill to stay home today without a change of reason, then it is not reasonable to expect Mr. Potter to refrain from liquidating Bailey's Building and Loan without a change of reason. Of course, FitzPatrick, in a sense, agrees with this. According to FitzPatrick, it is reasonable to expect Potter to refrain from liquidating Bailey's Building and Loan because it is reasonable to expect Potter to change his reasons. And it is reasonable to expect Potter to change his reasons because the very capacity that is required of Potter to change his reasons (*viz.*, the practice of being open to rival ideas) is a capacity that Potter both possess and employs in other domains of his life.

¹⁰⁸ It is likely that everyday Jill conceives of reasons to stay home (e.g., that she is tired, that she hates her job, that there will be a good a television show on this afternoon), but never conceives of these reasons as being sufficient to refrain from going to work.

But certainly in order for Potter to exercise this capacity, a reason in favor of doing so must at least present itself to Potter.¹⁰⁹ But as FitzPatrick describes the situation, Potter identifies the balance of reasons as weighing *against* exercising this capacity: the others living in Bedford Falls are just “socialists” or “liberal wackos,” he need not listen to their bleeding hearts. This means that in order for Potter to do what FitzPatrick believes is reasonably expected of him, he must act against his considered judgment that the residents of Bedford Falls are to be listened to when it comes to questions about morality. However, were Potter to act in such a way, he would be acting *akratically*, which is traditionally understood to be a form of practical irrationality. But, as Levy points out, it is only reasonable to expect of agents “those things they could ... do by way of some reasoning procedure.”¹¹⁰ But as a form of practical irrationality, *akrasia* results from a *breakdown* in reasoning, a failure to transform one’s considered judgment into action.¹¹¹ And so because fulfilling his epistemic obligations requires an irrational act on his part, it is not reasonable to expect Potter to do so.¹¹²

If Levy and I are correct, if it is not reasonable to expect a person to act *akratically*, then FitzPatrick’s weak condition on reasonable expectation is *too* weak. But notice that if reasonable expectation requires the realization (on the agent’s part) that

¹⁰⁹ I suppose that strictly speaking this is false. Potter can exercise this capacity accidentally or automatically. But certainly it is not reasonable to *expect* him to exercise the capacity accidentally; certainly what FitzPatrick has in mind is an *intentional* exercise of the capacity in question.

¹¹⁰ Neil Levy, “Culpable Ignorance and Moral Responsibility”, p. 739. The scope of this sentence is actually about the appropriateness of blaming agent’s for omissions they can arrive at by a reasoning procedure. In the context of Levy’s argument, and how it operates against FitzPatrick’s, the point is the same.

¹¹¹ There is, of course, a sense in which this line of reasoning is false. We can *accidentally* act against one’s best judgment, and in doing so not act *akratically* (traditionally, *akrasia* has been understood as *intentionally* acting against one’s best judgment).

¹¹² This does not mean that one should not try to *correct* Potter’s reasons.

reason dictates the course of action in question, then we have just transitioned to Rosen's strong condition on reasonable expectation.

3.4.2 A Possible Objection?

The reply advanced in 3.4.1 depends on the following plausible premise:

(K) It is not reasonable to expect a person to act contrary to her considered (best) judgment.

The plausibility of (K) rests on the assumption that acting against one's best judgment is a form of irrationality. That is to say, it is the *irrationality* of *akratic* action that makes it something unreasonable to expect.

A possible way to reject (K), and in doing so to defend FitzPatrick, is to reject the assumption that makes (K) plausible: reject that *akrasia* is necessarily irrational. In taking this route, one might wish to endorse a recent argument advanced by Nomy Arpaly¹¹³ according to which it is sometimes rational to act contrary to one's best judgment (or at least more rational than not). If her argument is sound, it might be thought, then we will have undermined the background assumption on which (K) rests. Without this background assumption, there is no standing presumption against it being reasonable to expect (under certain situations) *akratic* action.

I believe, however, that Arpaly's argument is somewhat deceptive. Even if sound, it can only be used to support FitzPatrick if one makes a certain equivocation, at least that is what I will argue.

¹¹³ Nomy Arpaly, "On Acting Rationally Against One's Best Judgment." It should be noted that Arpaly's essay (published in 2000) predates the Rosen/FitzPatrick/Levy debate.

Two things should be noted about Arpaly's argument. The first is that Arpaly distinguishes two ways of theorizing about rationality. The first way is "to see the ideal theory of rationality as providing us with a manual of sorts: follow these instructions and you will always make a rational decision."¹¹⁴ The second way is to see the ideal theory of rationality as determining from "a God's-eye view of a person's circumstances, beliefs, and motives" what actions would be considered rational for that person.¹¹⁵

The second thing to note is that Arpaly's argument proceeds from the perspective of *reasons internalism*. That is, Arpaly assumes from the outset that "one only has a reason to act in a certain way to the extent that the relevant course of action satisfies one's desires."¹¹⁶ With this notion of reason in hand, Arpaly understands rational action to involve "doing what one has overwhelming reasons to do," i.e. that which would satisfy (or not frustrate) most of one's desires.¹¹⁷

Now, it is certainly not Arpaly's position that it is *always* rational to act against one's best interest, or that such action is *often* rational. The claim she defends is much weaker: *sometimes* it is more rational to act against one's best judgment than in accordance with it.¹¹⁸ The cases Arpaly has in mind are quite specific in nature. They all involve individuals who are mistaken in their judgments about what they have most reason to do. Given reasons internalism, this means that the agents in question are

¹¹⁴ *Ibid*, p. 488

¹¹⁵ *Ibid*.

¹¹⁶ Arpaly, "On Acting Rationally," p. 492. This view contrasts with *reasons externalism*—the thesis that reasons are facts wholly independent of a persons desires or motivational set. According to reasons *externalism* *S* can have a reason to *p* regardless of the "contents" of her motivational set. To the reasons internalist, this is nonsense.

¹¹⁷ *Ibid*.

¹¹⁸ She admits that so acting will never be *fully* rational, just (in certain circumstances) more rational than not.

mistaken about what they *actually* most desire. It follows from the way she understands rational action that, since in such cases what the agent has “overwhelming reason to do” is that which is contrary to his or her best judgment, *akratic* action is more rational than *enkratic* action.

What makes Arpaly’s argument appear to be a promising way out for FitzPatrick is that its conclusion does not rely on “external” reasons, and thus it would appear to undermine the assumption that (K) rests on insofar as the irrationality of *akrasia* depends on a subjective assessment of reasons on the part of the agent. However, the distinction between subjective and objective rationality and that between external and internal reasons do not map on to one another. Rather, the distinction between subjective and objective rationality better maps onto the distinction that Arpaly draws between rationality understood as an agent’s manual and rationality understood from a God’s eye-view. The way in which I articulated subjective rationality above was in terms of an agent’s *awareness* of reasons; that is, reasons to which the agent has “access.” I articulated objective rationality in terms of the reasons that an agent actually *has*. The account was neutral with respect to the ontological status of such reasons. In other words, the objective/subjective distinction is an *epistemological* distinction, not an *ontological* one. As Arpaly indicates (and her argument requires), we can be unaware (whether through self-deception or simple lack of self-reflection) of many of our reasons even if reasons internalism is correct. Further, Arpaly is clear that it only makes sense to talk of acting rationally against one’s best interest in terms of a “God’s-eye view” of

rationality rather than a “rational agent’s manual.”¹¹⁹ That is, from the agent’s *point of view* in deliberation, it makes no sense to, for instance, judge that *I should really act against my best judgment here*, for then one would be forming a considered judgment and acting in accordance with it, thus not acting against one’s best judgment (not to mention how difficult it would be to discern from the first person perspective those moments when one *should* and *should not* act against one’s best judgment, for we can never really *know* if we have full awareness of relevant reasons).

Given these considerations, Arpaly’s thesis becomes relatively mundane. Despite the fact that she proceeds from a starting point of reasons internalism, we can understand Arpaly’s position as follows: because sometimes one will misjudge what one has most reason to do, and because rationality understood from the God’s eye view is doing what one has most reason to do, it will sometimes be rational (namely when one is wrong about what one has most reason to do) to act contrary to one’s best judgment. But when we understand that the God’s-eye view of rationality is just rationality considered *objectively*, then this is neither a controversial thesis, nor one at odds with the assumption underlying (K). That is, there is no disagreement that sometimes the objectively rational thing to do will be to act contrary to one’s best judgment (for the objectively rational thing to do is just to do that which one has most reason to do). The important question is this: is it reasonable to expect of a person that she do what it is objectively rational to do (i.e. rational from the God’s-eye view) even if doing so would

¹¹⁹ “As acting against one’s best judgment is something that no rational agent’s manual can advise...” *Ibid.* p. 490.

be subjectively irrational (i.e. violate the rational agent's manual)? This seems to be just too much to expect of a person.

4. THE SKEPTICAL ARGUMENT II: THE HARD SKEPTICAL UPSHOT

Section 3 defended Rosen's lemma, which is as follows:

(C3) S can be blameworthy for ϕ only if *akrasia* occurs somewhere in ϕ 's etiology (i.e. causal history).

The contrapositive of (C3) is:

(C3cp) If *akrasia* does not occur in ϕ 's etiology, then S cannot be blameworthy for ϕ .

In this section I will defend the following two claims:

(R) When an agent acts *akratically*, she does so as a result of a failure in her rational capacities;

and

(S) If an agent's action results from a failure in her rational capacities, then she is exempted from blame for that act.

It follows from (R) and (S) that:

(T) When an agent acts *akratically* she is exempted from blame for that act.

The conjunction of (C3cp) and (T) produce what I call the ***Hard Skeptical Upshot***:

Hard Skeptical Upshot: For any given agent S and any given act ϕ , either *akrasia* occurs in ϕ 's etiology or it does not. If it does not occur in ϕ 's etiology, then S cannot be blameworthy for ϕ (C3cp); but if *akrasia* does occur in ϕ 's etiology then S cannot be blameworthy for ϕ (T). Thus, for any given agent S , and any given act ϕ , S cannot be blameworthy for ϕ .

My defense of the **Hard Skeptical Upshot** shall be organized as follows: in section 4.1, I defend (R); in section 4.2 I defend (S). I save for section 5, what I take to be the importance of the Hard Skeptical Upshot.

4.1. *Akrasia* and Irrationality

In this section I defend (R): when an agent acts akratically she does so as a result of a failure in her rational capacities. The corresponding argument for this premise is as follows:

- 1.1 A rational agent is one who is able to act rationally. [Definition]
- 1.2 Rational action requires, in part, one to respond appropriately to reasons.
- 1.3 Responding appropriately to reasons requires, in part, doing what one judges oneself to have most reason to do (when one judges oneself to have most reason to do something).
- 1.4 So, if one fails to do what one judges oneself to have most reason to do, then one acts irrationally. [From 2.2, 2.3]
- 1.5 Irrational actions result from a failure in one's rational capacities.
- 1.6 So, when a rational agent fails to do what he judges to have most reason to do, he does so as a result of a failure in his rational capacities. [From 1.1, 1.4, 1.5]

In advancing this argument I shall focus on defending 1.2, 1.3, and 1.5. It should be noted that defending these premises relies on a controversial theory of rationality—a full defense of which lies outside the scope of this project. In its place I offer suggestions as to why such a theory may be correct.

4.1.1 Responding Appropriately to Reasons

The argument that follows relies on a condition I consider to be necessary for rational agency: what I call *responding appropriately to reasons*.¹²⁰ That this is a necessary condition for rational agency is controversial. While there is little agreement as to what it means to *respond* to a reason,¹²¹ there is even less agreement as to what constitutes responding *appropriately* to a reason. In other words, what distinguishes an appropriate from an inappropriate response is a vexing question. What follows are some suggestions as to how I think this distinction should be understood.

Before beginning, however, we need to make some preliminary distinctions. First, we need to distinguish *practical* rationality from *theoretical* rationality. As I shall understand these terms, the difference between practical and theoretical rationality lies primarily in their subject matter. The former concerns actions (and, possibly, certain practical attitudes, such as desires and intentions) while the latter concerns primarily cognitive attitudes, such as belief. When our concern is with what it is rational *to believe*, we find ourselves in the domain of theoretical rationality. On the other hand, when we find ourselves concerned with what it would be rational *to do*, we are in the domain of practical rationality.¹²² One can be theoretically rational and practically

¹²⁰ This has been called the “entailment thesis” by John Broome. See John Broome, “Does Rationality Consist in Responding Correctly to Reasons?” *Journal of Moral Philosophy*, 4 (2007), pp. 349-74. Broome characterizes his thesis as responding ‘correctly’ to reasons. I prefer the term ‘appropriately,’ though I am not sure how much hinges on this.

¹²¹ As I understand it, responding to a reason is a category broader than acting for a reason. That is, responding to a reason could include acting as well as belief formation.

¹²² Putting it this way can be misleading, and possibly obscure an interesting controversy in the literature. The controversy is this: do questions concerning what I *ought* to do fall within the domain of theoretical or practical rationality? That is, if I believe that I ought to jump off the edge of the Grand Canyon am I theoretically or practically irrational? While the subject of my belief is a practical matter, the fact that it is a belief, in my opinion, places it in the realm of theoretical rationality.

irrational or one can be theoretically irrational and practically rational.¹²³ The two domains are such that being irrational in one respect does not necessarily make one irrational in the other. In what follows I shall be primarily concerned with the practical side of rationality.

Next, we need to distinguish between *objective* rationality and *subjective* rationality. Some have argued that because rationality supervenes on the mind, the distinction between objective and subjective rationality is bogus.¹²⁴ From what I can tell, however, this view is the minority opinion.¹²⁵ I believe the distinction can be made by appealing to mental states of idealized agents. Objective rationality, as I shall understand it, concerns full knowledge and good reasoning. We can stipulate, then, that an agent who is objectively rational will always do the right thing for the right reason.¹²⁶ Subjective rationality, however, concerns only good reasoning. What distinguishes a subjectively rational agent from his objectively rational counterpart is not the reasons that each *has*—they have the same reasons in a certain sense—but rather the reasons that each has *access* to. That is, the objectively rational agent is fully aware of all the relevant reasons for acting, while the subjectively rational agent lacks this complete awareness. Because of this, the respective results of objective and subjective rationality

¹²³ Cf. Ralph Wedgwood, “Choosing Rationally and Choosing Correctly”, where it is argued that theoretical rationality is a necessary condition for practical rationality.

¹²⁴ Cf. John Broome, “Rationality” in *Blackwell Companion to Philosophy of Action*, edited by Timothy O’Connor and Constantine Sandis (Oxford: Wiley Blackwell, 2010): 285-292.

¹²⁵ I have seen the distinction made by, among others, Mark van Roojan (“Moral Rationalism and Arational Moralism,” *Ethics* 120 [2010], 495-525.), Christian Miller (“Gert on Subjective Practical Rationality” *Ethical Theory and Moral Practice* 11 [2008], 551-561), Sarah Stroud (“Moral Worth and Rationality as Acting on Good Reasons” *Philosophical Studies* 134 [2007], 449-456), Derek Parfit (“Reasons and Motivation” *Aristotelian Society Supplementary Volume* 71 (1997), 99-130) , and T.M. Scanlon (*What We Owe to Each Other*).

¹²⁶ “Right thing” and “right reason” need not be understood in moral terms.

will differ. The point of this distinction is to make clear that rationality, as I shall understand it, is a procedural concern. Its concern is not so much with the results as it is with the process or procedure whereby one gets to one's result.

I have stated that the difference between the objectively rational agent and her subjectively rational counterpart concerns the knowledge of the reasons that each respectively has. The kinds of reasons in question are *normative* reasons. If R is a normative reason for S to ϕ , then R is a consideration that counts in favor of S ϕ -ing. This can be contrasted with an *explanatory* reason. An explanatory reason is the reason for which an agent acts, which need not be a normative reason, i.e., it need not be a consideration in favor of acting. For example, I could eat a piece chalk because I am hungry. But my being hungry is not a consideration that counts in favor of eating chalk. Thus, my being hungry is not a normative reason for eating chalk. I shall stipulate that if R is a normative reason for S to ϕ , then R is a reason for S to ϕ regardless of whether or not S is aware of R .¹²⁷

We are now in a position to say a few things about what it is to respond appropriately to reasons. First, there is a knowledge constraint on responding appropriately to reasons. The reasons which are considered appropriate to respond to can only be those that the agent is *aware* of. If an agent is ignorant of a banana peel in her path and as a result steps on it and slips, we would not say that she failed to respond to the considerations that would have had her avoid the banana peel, for we can only expect

¹²⁷ This does not commit me to reasons externalism. Even if all reasons are internal, one can still have reasons that one is not aware of. Cf. Bernard Williams, "Internal and External Reasons," and Nomy Arpaly, "On Acting Rationally Against One's Best Judgment."

an agent to avoid a banana peel in her path if she is *aware* of there being a banana peel in her path.¹²⁸ Similarly, if she happens to avoid the banana peel in her path without realizing that there was in fact a banana peel in her path, we wouldn't say that she responded to the consideration that would have had her avoid the banana peel.¹²⁹ Thus, an agent can respond to a reason only if she is *aware* of this reason.

Second, the reasons that explain the response must be reasons that justify the response. For example, one's reason for acting must not only be an explanatory reason, it must also be a normative reason. Eating chalk is not an appropriate response to being hungry because being hungry does not justify eating chalk.¹³⁰ Now there is a difficulty when it comes to explaining exactly how a reason justifies an action. I won't try to solve this difficulty here, but I will take it that there is at least an intuitive understanding of when a reason justifies an action (as when my house being on fire justifies my jumping out the window) and when it does not (as in the chalk example). I take it that this is true not just of actions, but other sorts of possible responses to reasons. In order for a belief to be an appropriate response to reasons, the reasons which explain the belief must also *justify* the belief. One can, for example, believe that Muslims are evil because of the attacks of September 11, 2001. However, the latter is not a reason that justifies the former. In other words, the former is not an appropriate response to the latter, whereas the belief that airport security should be heightened may be an appropriate response.

¹²⁸ I take this condition to fall out of the distinction between objective and subjective rationality.

¹²⁹ Though we might say that in avoiding the banana peel, she did what she had most reason to do.

¹³⁰ I take it that one being hungry can be a normative reason. This is on the condition that one being hungry is different than simply the desire for food. While I don't argue it here, I don't take desires to be normative reasons.

That one's reason justifies one's action is not enough. Consider the following case:

Suitcase: Alan packs his suitcase everyday for a different reason. On Monday, Alan takes as his reason for packing his bag that Willie Nelson is a great singer. On Tuesday, Alan takes as his reason, that Plato is the greatest Western philosopher. Today, Alan takes as his reason that he is leaving town tomorrow. Alan is in fact leaving town tomorrow and this fact provides Alan with most reason to pack his suitcase.

I take it that on Monday and Tuesday Alan is not responding appropriately to reasons, for *that Willie Nelson is a great singer* or *that Plato is the greatest Western philosopher* (even if true!) is not a consideration in favor of packing one's suitcase. But does Alan respond appropriately to reason when he packs his bag today? After all, by hypothesis, the fact that he is leaving town tomorrow justifies his packing his suitcase today. Judging from Alan's history, however, it is likely to be a fluke that Alan's reason for packing his bag today justifies his action. Our final condition needs to rule out such cases of coincidence. What I propose is that there be some kind of counterfactual condition on the relationship between reason and response. That is, in order for Alan to be responding appropriately to reason today, it must be the case that were it not for the fact that he was leaving town tomorrow (better: were it not for the fact that he *took* himself to be leaving town tomorrow), he would not be packing his bags.¹³¹ But this is not enough, for this

¹³¹ That there might be a counterfactual condition on responding appropriately to reasons was first suggested to me by Linda Radzik in conversation. Robert Nozick takes a counterfactual condition to be required for knowledge. See his *Philosophical Explanations*. Broome also thinks there may be something like a counterfactual condition on responding appropriately to reasons. See note below.

condition might be met, but the relationship between Alan packing his bags when he takes himself to be leaving town tomorrow may be of the same *kind* as when he takes Willie Nelson to be a great singer as the reason for packing his bags. What we need then is this: were it not for the fact that he took himself to be leaving town tomorrow *and he understands that this counted as a consideration in favor of packing*, Alan would not have packed his bags today for the reason that tomorrow he is leaving town. In other words, in order for Alan to be responding appropriately to reason, he must understand that his leaving town tomorrow counts as a consideration in favor of packing his bags tonight and packs his bags *because of this*.

While I don't think that I've given all the necessary and sufficient conditions for responding correctly to reason, I take it that I've said enough at this point to give a picture as to what it means to respond appropriately to a reason.¹³²

4.1.2 Acting Against One's Best Judgment (Revisited)

I now move to my defense of premise 1.3: responding appropriately to reasons requires, in part, doing what one judges oneself to have most reason to do (when one judges oneself to have most reason to do something). While this will obviously be true when one's judgment (as to what one has most reason to do) is correct, it is not as obviously true when one's judgment is false. As we saw in section 3.4.2, Arpaly has

¹³² In "Does Rationality Consist in Responding Correctly To Reason," John Broome gives two necessary conditions for responding correctly to reason. His first condition is the following: one must ϕ whenever one has most reason to ϕ . This violates my knowledge condition. His second condition (to rule out coincidence, as I have) is: there needs to be an explanatory or counterfactual connection between reason and action. We saw that mere explanation (my second condition) or counterfactual connection (the first half of my third condition) were not enough to rule out coincidence. These are the only conditions Broome offers. However, like me, Broome is not interested in giving a complete account of this notion.

argued that when one's judgment as to what one has most reason to do is false, it may be the case that acting *against* that judgment is the rational response.¹³³ I shall presently consider a different aspect of that argument and conclude that it can't be the right way of thinking about things. This shall constitute my defense of premise 1.3.

When talking about what it is rational to do and what reason requires of us, it can be easy to fall victim to equivocation. When we ask such questions are we simply asking what it is that we have most reason to do? Or, alternatively, is our question about the reasoning process? My concern with practical rationality concerns the *process* whereby we go from reason to action. That is, questions about rationality are *procedural* questions. On the other hand, questions about what we have most reason to do are what we might call *substantive* questions.¹³⁴ It is generally taken that the answer to the substantive question about what we have most reason to do is not dependent on any particular judgment we happen to have. In other words, the fact that we judge ourselves to have most reason to do something will not change anything about what we have most reason to do. Alternatively: merely judging that she has most reason to ϕ does not in itself give *S* reason to ϕ .¹³⁵ So answering substantive questions about rationality requires answering only questions about what one has most reason to do.

Answering procedural questions about rationality is not as easy. These questions concern not what one *has* most reason to do, but rather what it is *reasonable* for a person

¹³³ I have in mind Nomy Arpaly, "On Acting Rationally against One's Best Judgment."

¹³⁴ I take the substantive/procedural distinction from T.M. Scanlon, *What We Owe to Each Other*, though I put it to different use.

¹³⁵ This issue sometimes goes under the name of "bootstrapping." Cf. Michael Bratman, *Intention, Plans, and Practical Reason*, (Palo Alto: Center for the Study of Language and Information, 1987); and John Broome, "Reasons."

to do given her situation. Facts about the agent's beliefs will play an important role in determining what it is reasonable for an agent to do. The question we want to answer is this: from a procedural point of view, how (if at all) does one's judgment about what one has most reason to do affect what it is most reasonable for that person to do?

As we have seen (in section 3.4.2), Arpaly has argued that under certain circumstances, rationality may not in fact require one to act according to one's best judgment. Under these circumstances the appropriate response to reason is not *always* to do that which you judge to have most reason to do. To illustrate her point, Arpaly appeals to an illustration involving a procrastinating student, Sam, who has found himself with only three weeks left to master difficult material before the final exam. Sam judges that what he has most reason to do is make himself a hermit for the next three weeks and do nothing but study. As it happens, however, Sam's "motivational set" is such that, were he to become a hermit for the next few weeks he would be miserable and would not be able to focus as well. Were Sam to isolate himself and become a hermit, many more of his desires would be frustrated than were he to refrain from isolating. In such a case, argues Arpaly, the rational thing for Sam to do is not to do that which he judges best—become a hermit—but rather refrain from isolating and remain balanced in his social life. Arpaly argues that were Sam to become a hermit, it is clear that we would judge him to be irrational.

While it may be rational for Sam to act against his judgment, Arpaly recognizes that Sam's doing so does not necessarily make his action rational. For in so doing, Sam may not be acting according to reasons, but rather may be succumbing to temptation,

laziness, or depression. But, Arpaly notes, acting against one's judgment need not take on this description. While the classic case of practical rationality takes on the form of deliberation followed by continent action, Arpaly argues that rational action need not go through the process of deliberation. Sam may be a poor deliberator, but he may yet be the kind of agent who is inclined to act only according to good reasons. "In failing to become a hermit," Arpaly states, "Sam might, unbeknownst to him, be acting for the very same reasons which he overlooked in his deliberation."¹³⁶ That is, just because Sam does not act according to his best judgment does not mean that he is not being responsive to reasons. The very considerations that give him most reason to refrain from isolating may be the very things that move him to do so.

While there is a certain air of plausibility to Arpaly's suggestions, I do not find them convincing. We might grant Arpaly that in not becoming a hermit, Sam is responding to the very reasons that speak in favor of doing so. However, as we saw in section 4.1.1, these two facts—that Sam has most reason to do so and that Sam did so because of these reasons—are not enough for Sam's act to be considered rational. The third condition on rational action—that the agent understand the reason for which she acts—is not met in the case of Sam. In fact, in refraining from isolating, even if he is doing so as a result of reasons that favor it, Sam does not take himself to be doing what he has most reason to do. In general, in acting against one's best judgment, this condition will never be met for, by hypothesis, one is not doing what one takes oneself to have most reason to do, and so in doing that thing one will not understand the reason

¹³⁶ Arpaly, "On Acting Rationally Against One's Best Judgment," p. 503.

under which one acts as justifying one's action. It is because of this that responding appropriately to reasons requires one to act according to one's best judgment (when one so judges). And when one fails to so act, one acts irrationally.

4.1.3 Rational Capacities

I shall now say something in defense of premise 1.5: irrational action results from a failure in one's rational capacities. While it may be somewhat trivial to say that irrational action cannot result from one's fully functioning rational capacities, the claim I am defending here is not so trivial. First, my claim posits a necessary condition for an action to be considered irrational, namely that it result from a failure in one's rational capacities. Second, that this condition is necessary is anything but obvious. What follows by way of defense is not a rigorous argument—that will have to wait for another time. Instead, what I offer is a set of considerations that I think speak in favor of the premise.

Much of the discussion up to this point has focused on *rational* action. In section 4.1.1 I argued that a necessary condition for an action to be considered rational is that it result from an appropriate response to reasons. It should be noted that the denial of this condition does not give us a sufficient condition for irrational action, but only a sufficient condition for an action to not be rational. This is the first thing I wish to point to. Irrationality is not merely the denial of rationality. There is a third option available: actions can be arational. I understand arational actions to be those actions to which the norms governing rationality fall silent. Examples of arational actions include my lifting my coffee cup up to my lips for a sip, my picking my pen up and putting it behind my

ear right ear, and my scratching my beard while I think about what sentence to write next. What unifies these actions as a class is that they are intentional actions to which reason neither speaks in favor nor disfavor. To say that scratching my beard while I think is rational would be strange. But it would be just as strange to say that it is irrational. The point is this: merely denying any condition we find to be necessary for rational action does not get us a sufficient condition for irrational action. This fact makes the task at hand all the more difficult. However, it also gives us a question to answer: what distinguishes irrational action from arational action?

It was stipulated that what makes an arational action distinct from either a rational or irrational action is that on the former, reason falls silent. I take this to mean two things. The first has already been mentioned: there are no considerations either in favor or against the act in question. The second concerns the perspective of the agent. If the act is arational, then the agent does not take the act to contribute toward any end she may have. Nor does the agent see reason as favoring the act. When an agent acts *arationally* she simply does it, sometimes for a reason, but sometimes for no reason at all.

Irrational action, then, can be understood to be an action which reason speaks *against*. We can say that an action (belief) is *substantively irrational* if it is something which one has reason not to do. We can say that an action is *procedurally irrational* just in case it results from an unsound reasoning process. Because *akrasia* is a form of procedural irrationality—it is not an appropriate response to reason—it must result from an unsound reasoning process. Now, if one's rational capacities are fully functioning at a

given time t , then at t one's reasoning process will be sound. If this is the case, then if at t one's reasoning process is not sound, it must be the case that at t one's rational capacities are not fully functioning. It follows that, because akrasia cannot result from a sound reasoning process, it must result from rational capacities that are not fully functioning, i.e. a failure in one's rational capacities. In other words, the following is true:

(R)When an agent acts *akratically* she does so as a result of a failure in her rational capacities.

4.2 A Failure in One's Rational Capacities

I now move on to my final premise: *if an agent's action results from a failure in her rational capacities, then she is exempted from blame for that act*. My defense for this premise is the following:

- 2.1 If an agent lacks relevant rational capacities at time t , she is exempted from blame for any act committed at time t .
- 2.2 If an agent's act results from a failure in her rational capacities, then the rational capacities that fail are among those relevant to ascriptions of blame.
- 2.3 If an agent's rational capacities fail at time t , then at time t the agent lacks those rational capacities.
- 2.4 So, if an agent's action results from a failure in her rational capacities, then she is exempted from blame for that act. (2.1, 2.2, 2.3)

This argument rests on Wallace's theory of exemptions, as articulated in section 2.

4.2.1 Exempting Conditions

Central to any theory of responsibility is an account of conditions whereby moral sanctions are deemed inappropriate. While it is not analytic to a theory of sanctions that it must have such conditions, in order for a theory of sanctions to fit the reality of our practical life—where we employ such notions as temporary insanity and duress—an adequate theory of sanctions must account for when and why sanctions are deemed inappropriate. The driving force behind the requirement that a theory of responsibility have as a component an account of exemptions is the intuition that under certain circumstances moral sanctions, such as blame and punishment, are considered inappropriate. What a proper account of responsibility needs to do, then, is either account for what it is about these circumstances that make sanctions inappropriate, or say something about why our pre-theoretic intuitions are mistaken.

As we saw in section 2, Wallace, like P.F. Strawson, distinguishes between two types of conditions whereby blame is deemed inappropriate. The first of these—excusing conditions—concerns the circumstances in which the act in question occurs. Sometimes our actions are only apparent wrongs, as when your being pushed on the bus was not due to my intentionally pushing you, but rather my being bumped by a person in front of me. While it may have been appropriate to sanction me were it the case that I intentionally pushed you, on finding out that the incident was the result of an accident, it would be inappropriate for you to continue resenting me.

My present concern is with the second type of conditions: exempting conditions. What distinguishes an exempting condition from an excusing condition is that whereas

the latter concerns the circumstances that give rise to the act in question, the former concerns features about the agent herself that make sanctions inappropriate. One way to think of this distinction is in terms of scope: an excusing condition is localized only to a particular act while an exempting condition is more global—it can apply to several (or even all) of an agent’s acts. An agent satisfies an exempting condition if she lacks one of what Wallace calls the powers of reflective self-control: (1) the power to grasp and apply moral reasons; and (2) the power to control or regulate one’s behavior by the light of reasons.¹³⁷ Because I discussed these powers in section 2, I will cut straight to the point: if Wallace’s theory of exemptions is correct, then premise 2.1 is true: *if an agent lacks relevant rational capacities at time t, she is exempted from blame for any act committed at t.*

4.2.2 Practical Irrationality and the Failure of Rational Capacities

The term ‘rational capacities’ denotes a heterogeneous group of powers that play many different roles. One such power is the ability to form consistent beliefs. One lacking this power will be prone to forming inconsistent beliefs such as that the world was made in six days and that the world was made in over a week.¹³⁸ Another power denoted by the term ‘rational capacities’ is that which translates choice into behavior. Yet another might include the ability to compute complex math equations. Naturally, not all of the powers which fall under the heading of a ‘rational capacity’ are directly relevant to the production of action; many of them, such as the power to form consistent

¹³⁷ Cf. *Responsibility and the Moral Sentiments*, p.157.

¹³⁸ This example comes, albeit in different context, from John Broome, “Reasons,”

beliefs or the power to draw implications, are only indirectly relevant to translating reasons into action.¹³⁹

Now, when I say that an act results from a failure in one's rational capacities, as I do in premise 3.2, I do not mean that it results *indirectly* from a failure in one's rational capacities, as when one's act is a result of an irrational belief. Rather, what I mean is that the act results directly from a failure in one's rational capacities. If an act results *directly* from a failure in one's rational capacities, the rational capacities in question must be those tied to practical rationality, for a failure in the powers of theoretical rationality can directly lead only to beliefs, not actions. But where must such a failure take place? We can say that generally, when an action results directly from a failure in one's rational capacities, it must be tied to the process that leads from reason to action. This process is one which I termed 'responding appropriately to reasons.' This process includes taking one's reason for acting as justifying one's action as well as comprehending how this reason justifies the action. In section 4.1.3 I suggested that practical irrationality is best understood as resulting from a failure in this process.

What is the relationship between the powers Wallace identifies as "the powers to regulate and control one's action by the light of moral reasons" and the process I call "responding appropriately to reasons"? I believe that the latter is involved in the former. That is, in order to regulate and control one's action by the light of moral reasons one must be able to respond appropriately to reasons. In order to deliberate and make a choice as a result of deliberation, one must be able identify reasons as considerations

¹³⁹ They are indirectly relevant insofar as they contribute to determining the candidate reasons for action.

(this is deliberation) and see how they justify a certain course of action (this is involved in making a choice). Similarly, in order to translate one's choice into action one must be moved into action by one's choice in virtue of the fact that it is what one judges ought to be done. But if this is the case, when the process of responding appropriately to reason fails, so do the capacities that are relevant to ascriptions of blame, since the former are an essential feature of the latter.

4.2.3 *Lacking the Ability to Exercise a Capacity*

I am now at the final stage of my argument. I have argued, to this point, that akrasia results from a failure in one's rational capacities and that those rational capacities that fail in akratic action are those that are relevant to ascriptions of blame. There will, however, be strong resistance to moving from this to the claim that akrasia satisfies an exemption from blame. An exemption from blame, it will be pointed out, occurs only when one *lacks* one of the relevant rational capacities. That a rational capacity fails does not mean that one lacks the rational capacity in question. In fact, in order for one's rational capacity to fail, one must *have* that rational capacity to begin with. If a gun fails to fire, this does not mean that one lacks a gun.

As intuitive as this line of thinking is, I will argue that it is wrong. When a rational capacity fails at time t one, in effect, lacks that rational capacity at that time. If this is correct, then it commits me to the rather unintuitive position that when one's gun

fails to fire at time t , one lacks at gun at that time.¹⁴⁰ I shall argue that this is not a problem.

As was mentioned in section 2 one of Wallace's main objectives in *Responsibility and the Moral Sentiments* is to defend the notion of responsibility against the incompatibilist's similar-structure argument (or as Wallace calls it, the generalization strategy). In defending against this argument, Wallace argues that were determinism true, it would not impair the powers of reflective self-control of all persons, all the time. The truth of determinism, argues Wallace, is perfectly compatible with our *possession* of such powers.

Wallace's defense against the generalization strategy is important for the present discussion because the move I make in arguing that a failure in one's rational capacities can be understood as an exempting condition is similar to the move the incompatibilist makes in arguing that being unable to do otherwise at t generalizes into an exemption. In order to properly defend my position, I need to say something about why Wallace's defense against the generalization strategy cannot be used also against me. In order to do that, however, it will help to say a few words about the similar structure argument again.

Wallace's account of exemptions rests on the following principle: "it is unreasonable to hold someone to a moral obligation if the person lacks the power to grasp and comply with the reasons that support the obligation"¹⁴¹ The incompatibilist argues that, if determinism is true, then for any given agent S 's act ϕ at time t such that

¹⁴⁰ This will be false if one owns other guns. To solve this we can formulate the position more precisely. In order to do so we need to give name to the gun. Call it 'G'. We can then say: If G fails to fire at time t , then at t G is not a gun.

¹⁴¹ Wallace, *Responsibility and the Moral Sentiments*, p. 184.

in ϕ -ing S fails to fulfill her moral obligation, S cannot refrain from ϕ -ing at time t and so cannot at t comply with the reasons that support her obligation to refrain from ϕ -ing. In other words, according to the incompatibilist, the truth of determinism would, in effect, “deprive S of the ability to exercise the general powers in the particular circumstances of action.”¹⁴²

Wallace argues that this strategy rests on an equivocation. The principle underlying exemptions requires only the general possession of the relevant rational capacities, it does not require the more specific ability to *exercise* those capacities at any given time.¹⁴³ In defense of this position, Wallace appeals to an analogy involving language. Intuitively, we would judge it unfair to demand that a newly arrived non-English speaking refugee speak and understand English like a native speaker. The reason it would be unfair to demand this of the refugee, says Wallace, is that “the agent lacks the basic power to do the sort of thing that we demand of him.”¹⁴⁴ That is, according to Wallace, it is because the refugee lacks the general power to speak English like a native speaker—and not the fact that the refugee lacks the opportunity to exercise any abilities she may possess—that makes demanding she speak English unfair. That is, what is of concern to a theory of exemptions is just the *possession* of the relevant capacities, not the opportunity to exercise those capacities in any given situation.

Setting this aside for the moment, I would like to turn back to my proposal: if an agent’s rational capacities fail at time t , then at t she lacked those rational capacities. I

¹⁴² *Ibid.*, p. 185.

¹⁴³ For this argument, see *ibid.*, pp. 185-186.

¹⁴⁴ *Ibid.*, 161.

would imagine that the line Wallace would take against this would be similar to that he takes against the generalization strategy, for as we shall see, my proposal has several structural similarities with the incompatibilist's generalization strategy. Before getting to this, I need to say something about why I think my proposal is correct.

What is the difference between possessing a capacity to do a certain thing at a certain time and lacking that capacity to do that thing at that time? It's clear that a person has a capacity to do a certain thing at a certain time if she *does* that thing at that time. What if she doesn't do that thing at that time? Does she still have that capacity? So long as the person is *able* to do that thing at that time, she has the capacity to do that thing. That Albert Pujols fails to hit a home run during a game does not mean that he lacks the capacity to do so. But what if a capacity to do a certain thing at a certain time *fails*? Does the person still *possess* that capacity to do that thing at that time? If Pujols' capacity to hit a home run at *t* *fails him*, does he still possess the capacity to hit a home run? Imagine that Pujols' capacity to hit a home run requires the correct operation of certain muscles in his biceps. Imagine that these muscles spasm in such a way that they become inoperative at *t*. Does Pujols still have the capacity to hit a home run at *t*? Even if the spasm lasts for only a moment, it would still seem that at least at the time of the spasm, Pujols lacks the capacity to hit a home run, for he lacks the requisite operation of muscles that home-run hitting requires. What's different about this example and, for instance, one involving Pujols breaking his arm and losing the capacity to hit a home run, is that the latter involves an external impediment to Pujols' capacity to hit a home run while the former involves something internal to the capacity—the failure of the

capacity itself—that makes it such that he lacks the capacity. That is, what’s important about the Pujols example is that it involves a failure of a capacity which intuitively results in the loss of a capacity itself. Now, I don’t believe that there is anything special about this capacity that would make it different in this respect to capacities in general. And if this is so, if we can generalize the Pujols example to capacities in general, then we can say that when one’s rational capacities fail at t , one lacks those capacities at that time.

Wallace is likely to respond to this line of reasoning by arguing that I am making the same equivocation that the incompatibilist makes in advancing the generalization strategy. When Albert Pujols breaks his arm, he only lacks the ability to *exercise* his capacity to hit a homerun. Similarly, when Pujols’ suffers from a muscle spasm, he merely lacks the ability to *exercise* his capacity to hit a homerun at t , but he still possesses the general capacity to do so. If this is the case, when an agent’s rational capacities fail in the sense in question, at most she will lack only the localized ability to *exercise* these capacities, not lack the possession of the capacity in general. After all, in order for a capacity to *fail* one must have it in the first place.

However, if we reflect on some of the cases Wallace discusses under the heading of exemptions, we’ll see that distinguishing between lacking the general capacity and lacking the ability to *exercise* a capacity is not so easy to distinguish. First, consider the case of the refugee. According to Wallace, it would be unfair to demand the refugee to speak English like a native because she lacks the general capacity to speak English. Note that this cannot mean the foreigner is constitutionally incapable of speaking English.

Surely if she studied intensely enough she could learn. In some general sense of the word, she has the *ability* to speak English, she just lacks the requisite tools to do so. Now consider the case of the hypnotized individual who, according to Wallace, satisfies an exempting condition. If the reason the hypnotized individual satisfies an exemption is that she lacks the ability to reflect on her reasons, it can't be that she *completely* lacks this ability, for it is certainly something she can do when not hypnotized. Rather, her hypnotism acts as an impediment; it is only during the duration of her hypnotic spell that this ability is lacked.¹⁴⁵

What I am trying to get at is this: if it is unfair to expect Pujols to hit a homerun while his arm is broken, this is because he lacks the capacity to do so. When his arm heals, Pujols may regain the capacity to hit a home run, but a temporary incapacity is an incapacity no-less. Similarly with the muscle spasm: that the spasm is only temporary does not make it any less of an incapacity.

But even if we accept that Pujols' muscle spasm results only in the lack of an ability to *exercise* this capacity—and not a lack of the capacity itself—would it not still be unfair to expect him to exercise it? Imagine that I have a television set that has the capacity to transmit cable broadcast. However, because I failed to pay my cable bill, the cable company has disconnected my cable. Presumably, my television still has the capacity to transmit cable signals. However, because my cable has been shut off, it now lacks the ability to exercise this capacity. Would it be appropriate of me to expect my television to transmit cable broadcasts because it still possesses the capacity to do so?

¹⁴⁵ This impediment is, of course, an external impediment.

Certainly not. So how, if a person lacks the ability to exercise a capacity at a certain time, can it be fair to demand that she do that thing at that time?¹⁴⁶ It would seem that even if the failure of one's rational capacity is understood to only result in lacking the ability to *exercise* the capacity at that time we get the same morally relevant result as we do if she simply *lacks* that capacity.

If the line of reasoning in this section is correct, then premise 2.3 of the argument is true: if an agent's rational capacities fail at time *t*, then at time *t* the agent lacks those rational capacities. And it follows, from my defense of premises 2.2 and 2.1, that the following is true:

(S) If an agent's action results from a failure in her rational capacities, then she is exempted from blame for that act.

4.3 The Hard Skeptical Upshot

In Section 4.1 I defended the following claim:

(R) When an agent acts *akratically*, she does so as a result of a failure in her rational capacities.

This claim, together with (S) entails the following:

(T) When an agent acts *akratically* she is exempted from blame for that act.

In Section 3 I defended the following claim, which is the contrapositive of a lemma that occurs in Rosen's argument:

¹⁴⁶ It should be noted that if this reasoning is correct, then Wallace's defense against the generalization strategy fails, for there is no important distinction—in terms of whether it is fair to make certain demands of them—between lacking the ability to exercise a capacity at time *t* and lacking that capacity (simpliciter) at that time.

(C3cp) If *akrasia* does not occur in ϕ 's etiology, then S cannot be blameworthy for ϕ .

(T) and (C3cp) give us what I call **The Hard Skeptical Upshot**:

Hard Skeptical Upshot: For any given agent S and any given act ϕ , either *akrasia* occurs in ϕ 's etiology or it does not. If it does not occur in ϕ 's etiology, then S cannot be blameworthy for ϕ (C3cp); but if *akrasia* does occur in ϕ 's etiology then S cannot be blameworthy for ϕ (T). Thus, for any given agent S and any given act ϕ , S cannot be blameworthy for ϕ .

As we saw, the variety of skepticism that we find in Gideon Rosen's "Skepticism about Moral Responsibility" is much softer than this. According to Rosen, blameworthiness is perfectly possible and perfectly real. His skeptical upshot is epistemic in nature, and results from the fact that identifying instances of genuine *akrasia* is simply something that we cannot do with confidence.

Interestingly, Rosen notes the following: "If there were an independent argument for the thesis that akratic action is never culpable, then we would indeed have a general argument for the impossibility of responsibility."¹⁴⁷ This is just what I have provided.

¹⁴⁷ Rosen, "Skepticism about Moral Responsibility" p. 308.

5. CONCLUSION: THE MORAL UPSHOT

I shall conclude, in this section, with some reflections on what I take to be the moral upshot of the skeptical position advanced in this essay and indicate where I think future work needs to be done. Before doing that, however, I'd like to briefly summarize the argument that I've advanced.

5.1 Summary

My argument can be understood as proceeding in three stages. The first stage is best conceived of as laying the foundation for the latter, more substantive part of the argument. This stage involved setting out the position of Rational Capacity Compatibilism and the corresponding theory of exempting and excusing conditions. The second stage of the argument involved defending Gideon Rosen claim in "Skepticism about Moral Responsibility," that an agent cannot be blameworthy for an act if it is not the causal upshot of an akratic episode. The final stage of the argument defended that claim that *akrasia* satisfies an exempting condition from blame. This, in turn, involved defending the claim that *akrasia* necessarily results from a failure in one's rational capacities and that a failure in one's rational capacities amounts to lacking that rational capacity at that time. Given that one satisfies an exempting condition when one lacks said rational capacities, it follows that *akratic* action cannot be blameworthy. Thus, we find that whether or not an action results from *akrasia*, an agent cannot be blameworthy for that act.

5.2 Reflections

Unlike most skeptical arguments about blameworthiness, the one advanced in this essay is silent on many controversial metaphysical principles, such as free will and determinism. The hard skeptical upshot I arrive at in section 4 relies only on the nature of rationality and our rational capacities and the internal structure of blameworthiness itself. Granted, some of the arguments advanced there will be controversial, but the important thing to note is that unlike the notions of free will or determinism, the notions of rationality and rational capacities are *internal* to the notion of blameworthiness itself. This, I think, is important in that it indicates that the practice of blame is *internally unstable*. That is, my criticisms are internal to the practice of blame itself.¹⁴⁸

As I mentioned in section 1, my argument may be taken by some merely as a *reductio* of Wallace's theory of moral responsibility. At the moment, I don't have a way to reject this move. My argument is also not effective against the economy of threats approach mentioned in section 2. However, as a great number of people find this approach unappealing, I don't consider it a serious contender.

In section 2, I sketched an argument advanced by Strawson which Wallace calls "the pragmatic argument." According to this argument, even if we had reason to abandon our reactive attitudes (in the case of the original argument this reason was the truth of determinism) we should not, for our reactive attitudes are an important and crucial component of interpersonal lives. Without the rich interpersonal relationships that our reactive attitudes afford us, we will grow to view each other as mere objects of

¹⁴⁸ I get the idea of an internal criticism from Wallace, *Responsibility and the Moral Sentiments*, p. 101.

social policy. As such, even if the practice of blame is internally unstable, it is better to keep it in place than live in such a world as that.

I imagine that Strawson's pragmatic argument captures many intuitions. A world where we are mere objects of social policy sounds like a frightening dystopic vision. However, I find the worry to be a bit overblown, and in the case of the argument I've advanced in this essay, not applicable. The first thing to note is that the argument I have advanced has said nothing about praise, love, hate, dislike, and other such attitudes. My argument is *only* in regards to blameworthiness and does not work against praiseworthiness or what we might call loveliness.¹⁴⁹ Thus, the only reactive attitudes we need give up are those associated with the practice of blame. It is with a reflection on this last point that I'd like to end.

What would a world without blame look like? Some might think that without blame, individuals would have little incentive to act morally. This, I think, is nonsense. A world without blame is not a world without reasons or rationality. Moral reasons would still have the same weight, and assuming that we would still be in possession of our rational capacities (and assuming that we in fact *do* have rational capacities [if we don't we couldn't be blameworthy anyway]), we would have the exact same reasons to act morally as we do in a world without blame. As a good Kantian might say, incentives to avoid blame do not a moral reason make.

¹⁴⁹ This indicates the possibility of a vindication, of sorts, of Susan Wolf's asymmetric theory of responsibility. However, I would consider the asymmetry that results from my argument a *hard* asymmetry (blameworthiness is *impossible*) whereas Wolf's asymmetry is a *soft* asymmetry (blameworthiness *is* possible in indeterministic settings).

Finally, a world without blame is not a world without morality. Actions would still be deemed right or wrong, good or bad, moral or immoral. There would still be reason to do right and reason to refrain from doing wrong. What would change is the attitude we take toward another when one of us goes wrong. The contours of morality without blame is, unfortunately, something I don't have room to explore—that would likely take another hundred pages—but is something I hope to explore in later writings.

REFERENCES

- Arpaly, Nomy. "On Acting Rationally Against One's Best Judgment." *Ethics* 110, No. 3 (2000): 488-513.
- Bratman, Michael. *Intentions, Plans and Practical Reason*. Palo Alto, CA : Center for the Study of Language and Information, 1987.
- Broome, John. "Reasons." In *Reason and Value, Themes from the Moral Philosophy of Joseph Raz*. Edited by R. Jay Wallace, et al, 28-55. Oxford: Oxford University Press, 1994.
- "Does Rationality Consist in Responding Correctly to Reasons?" *Journal of Moral Philosophy*, 4 (2007): 349-374.
- "Rationality." In *Blackwell Companion to Philosophy of Action*. Edited by Timothy O'Connor and Constantine Sandis, 285-292. Oxford: Wiley Blackwell, 2010.
- Doris, John M and Dominic Murphy. "From My Lai to Abu Ghraib: The Moral Psychology of Atrocity." *Midwest Studies in Philosophy*, XXVI (2007): 25-55.
- Fischer, John Martin and Mark Ravizza. *Responsibility and Control: A Theory of Moral Responsibility*. New York: Cambridge University Press, 1998.
- FitzPatrick, William. "Moral Responsibility and Normative Ignorance: Answering a New Skeptical Challenge." *Ethics* 118, No. 4 (July 2008): 589-613.
- Frankfurt, Harry. *The Importance of What We Care About*. New York: Cambridge University Press, 1988.
- Holton, Richard. *Willing, Wanting, Waiting*. Oxford: Oxford University Press, 2009.

Levy, Neil. "Culpable Ignorance and Moral Responsibility: A Reply to FitzPatrick."

Ethics 119, No. 4 (2009): 729-741.

Miller, Christian. "Gert on Subjective Practical Rationality" *Ethical Theory and Moral*

Practice 11 (2008): 551-561

Parfit, Derek. "Reasons and Motivation." *Proceedings of the Aristotelian Society*, suppl.

71 (1997): 99-130.

Rosen, Gideon. "The Case for Incompatibilism." *Philosophy and Phenomenological*

Research 64, No. 3 (2002): 699-706.

--- "Skepticism about Moral Responsibility." *Philosophical Perspectives* 18 (2004): 295-

313.

Scanlon, T.M. *What We Owe to Each Other*. Cambridge, MA.: Harvard University

Press, 1998.

Stalnaker, Robert. "A Theory of Conditionals." In *Studies in Logical Theory*. Edited by

Nicholas Rescher, 98-112. American Philosophical Quarterly Monograph Series, vol 2. Oxford: Blackwell, 1968.

Strawson, Peter. "Freedom and Resentment." *Proceedings of the British Academy* 48

(1962): 1-25. Reprinted in *Free Will*, second edition. Edited by Gary Watson, 72-93. New York: Oxford University Press, 2003.

Stroud, Sarah. "Moral Worth and Rationality as Acting on Good Reasons."

Philosophical Studies 134 (2007): 449-456.

van Roojan, Mark. "Moral Rationalism and Rational Amoralism." *Ethics* 120 (2010):

495-525.

Wallace, R. Jay. *Responsibility and the Moral Sentiments*. Cambridge, MA: Harvard University Press, 1994.

Watson, Gary. "Free Agency." In *Free Will*, second edition. Edited by Gary Watson, 337-351. New York: Oxford University Press, 2003.

Wedgwood, Ralph. "Choosing Rationally and Choosing Correctly." In *Weakness of Will and Practical Rationality*. Edited by Sarah Stroud and Christine Tappolet, 201-229. Oxford: Oxford University Press, 2003.

Williams, Bernard. *Moral Luck*. Cambridge, UK: Cambridge University Press, 1981.

Wolf, Susan. *Freedom within Reason*. New York: Oxford University Press, 1990.

VITA

Zachary Charles Bachman received his Bachelor of Arts degree in philosophy from University of California, San Diego in 2009, earning *summa cum laude* and department honors (with high distinction) for his undergraduate thesis, “The Will and Salvation: Making Sense of Schopenhauer’s Soteriological Doctrine.” He received his Master of Arts degree in philosophy from Texas A&M in May 2011. Zachary Bachman’s research interests include philosophy of action, moral psychology, and ethics, with a particular interest in moral responsibility, blame, and practical reasoning. In the fall of 2011 Zachary will join the PhD program at University of California, Riverside. He may be reached by email at zac.bachman@gmail.com or by mail via the following address: Zachary Bachman, c/o UC Riverside Philosophy Department, HMNSS Building Room 1604, 900 University Avenue, Riverside, California, 92521.