

UC Riverside

UC Riverside Electronic Theses and Dissertations

Title

"What Bestiality of Thought": A Nietzschean Critique of Guilt and Punishment and the Economics of Suffering and Cruelty

Permalink

<https://escholarship.org/uc/item/36g1h45c>

Author

Johnson, Mark Cunningham

Publication Date

2018

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
RIVERSIDE

“What Bestiality of Thought”:
A Nietzschean Critique of Guilt and Punishment
and the Economics of Suffering and Cruelty

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

Doctor of Philosophy

in

Philosophy

by

Mark Cunningham Johnson

June 2018

Dissertation Committee:
Dr. Maudemarie Clark, Co-Chairperson
Dr. John Martin Fischer, Co-Chairperson
Dr. Andrews Reath
Dr. Michael Nelson
Dr. Gavin Lawrence

Copyright by
Mark Cunningham Johnson
2018

The Dissertation of Mark Cunningham Johnson is approved:

Committee Co-Chairperson

Committee Co-Chairperson

University of California, Riverside

Acknowledgments

My sincerest gratitude goes to the members of my committee: Maudemarie Clark, John Martin Fischer, Gavin Lawrence, Michael Nelson, and Andrews Reath. It was my great fortune to arrive at UCR the same year as Maude, it was through her expertise and patience that I have learned anything that I know about Nietzsche. The rest is not her fault. John is the undisputed master when it comes to guiding graduate students. He motivated me to begin writing when I was in a slump and helped me produce the chapters on Hershenov and Velleman. Gavin was an always enthusiastic conversational partner when it came to Nietzsche. His comments on several papers helped convince me that I might have a positive account to offer and might amount to more than a nay-sayer. Michael managed to combine unceasing encouragement with relentlessly challenging criticism and was on the committee that led to what is now the chapter on Velleman. Andy taught the first seminar I attended at UCR and the only reason I hesitate to call it the best is because I went on to take his other seminars. Andy's seminars are the Platonic form for graduate seminars and are the reason I know anything about ethics. I think it only appropriate to say that, while I owe you all a debt that I can never repay, I will feel no guilt in taking what you so generously offered...

If a fellow graduate student could be a member of a dissertation committee, then Avery Snellson would have been my sixth member. It was, once again, my good fortune that he arrived at UCR shortly after I did, came from a similar background, and was

interested in the same topic. I benefitted greatly from all our conversations and owe another debt that I can't repay.

Ironically, David Kelsey, a theologian, was the first person to see promise in my graduate-level philosophical writing and give me much appreciated encouragement when I needed it the most. Matthias Risse taught the first Nietzsche seminar I took, while a student at Yale Divinity School. I hope something in this dissertation makes up for the awful mess he had to read years ago.

My undergraduate philosophy professors inspired and encouraged me always: Malcolm Reid, David Aiken, Grady Spires, and Pat Fitzgerald, thank you all.

Over the years there were numerous attendees of the Nietzsche seminars (otherwise known as the "Maude Squad") whose contributions are not to be overlooked: Justin Coates, Joshua Hollowell, Justin White, Monique Wonderly, Daisy LaForce, Yvonne Tam, Will Swanson, Stephanie Allen, Max McCoy, Micaela Quintana, Rebecca Harrison, and David Beglin.

In addition, I've benefitted greatly, whether in specific or general philosophical terms, from conversations with: Howie Wettstein, William Bracken, Mark Wrathall, Gary Watson, Herb Morris, Harry Frankfurt, Sheldon Smith, Andrew Hsu, Larry Wright, Calvin Normore, John Carriero, Mia Wood, Nathan Lowe, Fahme El'Amin, and many others.

Lastly, my deepest appreciation goes to all of the “office mates” I’ve had over the years writing at the Edgemar Plaza and Peet’s Coffee shop (RIP). You’re an amazing bunch of weirdos.

Dedicated to
Pamela Hieronymi
for Love and Patience
Above and Beyond...

ABSTRACT OF THE DISSERTATION

“What Bestiality of Thought”:
A Nietzschean Critique of Guilt and Punishment
and the Economics of Suffering and Cruelty

by

Mark Cunningham Johnson

Doctor of Philosophy, Graduate Program in Philosophy
University of California, Riverside, June 2018
Dr. Maudemarie Clark & Dr. John Martin Fischer, Co-Chairpersons

ABSTRACT:

In *Twilight of the Idols*, Nietzsche proposes eliminating the concepts of guilt and punishment from our social institutions and sanctions. I argue that doing so is a plausible project that entails no negative repercussions to our ethical and social lives, and is desirable from a moral standpoint. Nietzsche’s genealogy of moral guilt shows him to be exclusively concerned with the feeling of “pervasive guilt” produced by the Judeo-Christian tradition. Pervasive guilt as a moral-religious interpretation of human suffering and feelings of personal inadequacy provides a pretext to satisfy cruelty through self-punishment for moral wrongdoing. This feeling survives declining theistic belief and remains a feature of secular cultures and moralities. Pervasive guilt is not a moral

emotion. It encourages self-cruelty as a means to feeling power and focusses attention on feelings of personal failure, rather than on the person wronged. Eliminating guilt from human psychology restores the conscience to the function of guiding ethical behavior through responsiveness to others, rather than self-punishment. Nietzsche's genealogy reveals punishment's roots in human cruelty and the desire to retaliate against the cause of injury. The concept of punishment emerges from primitive economic thinking that posits equivalences between injuries received and the injurer's suffering as compensation provided to the injured party. As a community's response to crime, punishment has had different expected results, but remains essentially a justification for inflicting suffering on the wrongdoer, becoming the unquestioned means for addressing crime. But punishment only anesthetizes the feeling of being wronged and replaces it with the pleasure of making suffer. Recognizing punishment as an interpretation allows a distinction between sanctions and punishment. Sanctions for crime can be reinterpreted as having the aim of protecting the community without punitive intent. Thus, eliminating punishment does not entail leaving crime unaddressed. Eliminating punishment aims at resisting the distorting demands of the victim's resentment, it avoids anesthetizing moral protest through violent affect, and aims to direct the criminal's attention away from their own suffering and towards recognition of the wrong done to the victim.

Table of Contents

Preface and Introduction	1
Eliminating Punishment	2
Eliminating Guilt	5
Nietzsche’s Project	6
Thesis	7
Chapter Outline	16
Chapter One: Debt Consciousness and the Creditor-Debtor Relationship	21
“Whence has this age-old, deeply-rooted, perhaps now no longer eradicable idea taken its power — the idea of an equivalence between injury and pain? I have already given it away: in the contractual relationship between creditor and debtor...”	21
Introduction	21
1.1 The Creditor-Debtor Relationship: Individual to Individual	23
1.2 Problems with the Creditor-Debtor Relationship	32
1.2.1 Memory & Promising	32
1.2.2 The Conscience	36
1.2.3 The Oldest Relationship Between Persons	42
1.2.4 Older Than any Societal Associations or Organizational Forms	47
1.2.5 Human Cruelty	53
1.3 The Creditor-Debtor Relationship: Community and Individual	61
1.3.1 Impact on Debt Consciousness	72
1.4 The Creditor-Debtor Relationship: Ancestors to Community	75
1.4.1 Implications for Debt Consciousness	79
Chapter Two: The Bad Conscience and Guilt	88
“Oh, this insane sad beast man! What ideas occur to it, what anti-nature, what paroxysms of nonsense; what bestiality of idea immediately breaks forth when it is hindered only a little from being a beast of deed! ...”	88
Introduction	88
2.1 The Influence of Debt Consciousness	94
2.2 The State: Conquest and Subjugation	98
2.2.1 Internalization of the Instincts	100
2.2.2 The Will-To-Power	102

2.2.3 The Animal Bad Conscience	104
2.3 The Instinct to Obedience and the Instinct for Freedom	106
2.4 Necessary Conditions for the Bad Conscience	110
2.5 The Development of the Bad Conscience	117
2.5.1 Noble Morality: “Good and Bad”	119
2.5.2 The Source of Ressentiment	120
2.6 The Moralization of Guilt	131
2.6.1 The Architect of Guilt: The Ascetic Priest	134
2.6.2 The Ascetic Interpretation of the Animal Bad Conscience	136
2.7 The Atonement as Temporary Remedy	141
2.8 The Implications of Moralization: Guilt Survives Atheism	150
2.9 The Bad Conscience: A Sickness Like Pregnancy	154

Chapter 3: Historical Instances: Post-Exilic Judaism and the Ancient Greeks **159**

“Then devise for me the love that bears not only all punishment but also all guilt! Then devise for me the justice that acquits everyone except the one who judges!”	159
Introduction	159
3.1 Post-Exilic Judaism and the Bad Conscience	161
3.2 The Slave Revolt in Morality	170
3.3 The Greeks and the Bad Conscience (or lack thereof...)	174
3.4 The Conscience and the Bad Conscience: “Episodic” vs. Pervasive Guilt	185

Chapter Four: Nietzsche on Punishment and Justice **190**

“I do not like your cold justice; and from the eye of your judges there gazes always the executioner and his cold steel.”	190
Introduction	190
4.1 Punishment as Interpretation	191
4.2 Punishment in the Creditor-Debtor Relationship	196
4.3 Punishment Does Not Awaken Moral Guilt	203
4.4 Punishment and Justice	209
4.5 Punishment as an Interpretation of Sanctions	225

Chapter Five: Equilibrium Through Revenge: **233**

Hershenov's Debt Theory of Punishment **233**

“... that humanity might be redeemed from revenge: that is for me the bridge to the highest hope and a rainbow after lasting storms.”	233
Preliminary: A Nietzschean Critique of Punishment	233
5.1 Introduction to Hershenov	237
5.1.1 Preliminaries: The Failure of Retributive Theories	239
5.1.2 Hershenov’s Debt Theory of Punishment	240
5.2 Critique of Hershenov	251
5.3 Crime and Motivation	252
5.3.1 Case 1: The Malicious Driver	253
5.3.2 Case 2: The Thoughtless Driver	256
5.3.3 Case 3: The Risk-Taking Driver	258
5.3.4 Case 4: The Getaway Driver	261
5.4 Eye for Eye: Immorality for Immorality	268
5.5 Justifiable Hatred and the Desire to Cause Harm	274
5.5.1 The Twisted Paramedic	278
5.5.2 Over-Eager Self-Defense	278
5.6 The Malicious Driver: The Irrelevance of Symmetrical Intentions	280
5.7 Jailhouse Morality and The Irrelevance of Hypocrisy	282
5.8 Debt Revised: Recognition of Moral Standing	285
5.9 The Sanction and Rehabilitation Model of Criminal Justice	287

Chapter Six: Undeserved Survival: Velleman on Survivor Guilt 292

Guilt. – Although the shrewdest judges of the witches and even the witches themselves were convinced of the guilt of witchcraft, this guilt still did not exist.	292
This is true of all guilt.	292
Preliminary: Nietzsche on Guilt	292
6.1 Introduction to Velleman & Survivor Guilt	296
6.1.1 The Set-Up: Moral Guilt vs. Survivor Guilt	297
6.2 Velleman’s Project	299
6.2.1 Moral Guilt: Loss of Trust & Normative Vulnerability	300
6.2.2 Normative Vulnerability & Survivor Guilt	304
6.3 Initial Worries About Envy and Resentment	311
6.4 Critique of Velleman: Desert, Justification & Normative Standards	314
6.4.1 The Chronic Invalid’s Envious-Resentment of the Healthy Person	315
6.5 Undeserved Goods: Having a Good One Should Not Have	318
6.6 Survivor Guilt Revised?	327
6.7 Lacking Normative Standards & Lacking Justification: Where It All Went Wrong	331
6.8 Concluding Observations on Velleman’s View	333

6.8 Normative Vulnerability vs. Relational Vulnerability: A Nietzschean Counter-Proposal	337
Conclusion: The Elimination of Guilt, Punishment, and the Economics of Suffering	349
Bibliography	379

Note on Texts and Citations

All references to Nietzsche's works are included within the text and follow the standard practice of employing the following abbreviations:

A	<i>The Antichrist</i>
BGE	<i>Beyond Good and Evil</i>
D	<i>Daybreak</i>
EH	<i>Ecce Homo</i>
GM	<i>On the Genealogy of Morality</i>
GS	<i>The Gay Science</i>
HH	<i>Human, All too Human</i>
TI	<i>Twilight of the Idols</i>
Z	<i>Thus Spake Zarathustra</i>

Preface and Introduction

“Thus, however, I counsel you my friends: Mistrust all in whom the drive to punish is powerful! They are folk of poor breed and stock; out of their faces leer the executioner and the bloodhound.”¹

This aim of this dissertation is to present and explain Nietzsche’s views on the origins, development, and nature of the concepts of guilt and punishment, which are primarily found in the “Second Treatise” of his *On the Genealogy of Morality* (GM). However, the puzzle that inspired and motivated this effort, and provides both its organizational thread and a hint at a solution, came from another text. In *Twilight of the Idols* “The Four Great Errors, 7,” (TI) Nietzsche claims that “we immoralists especially are trying with all our might to remove the concept of guilt and the concept of punishment from the world and to purge psychology, history, nature, the social institutions and sanctions...” What follows in this dissertation began primarily as an attempt to understand why Nietzsche believes that guilt and punishment ought to be eliminated and what would be entailed by their elimination. Ultimately, I will defend Nietzsche’s claims as plausible and as demonstrating a deep concern and engagement with morality.²

¹ Z:II:7

² The term morality is problematic in Nietzsche’s work and is potentially the source of much confusion. When reading Nietzsche it is helpful to keep mind the distinction between broader and narrower interpretations of morality. Nietzsche’s objections to “morality” are objections to overly narrow interpretations of it, such as are found in the Judeo-Christian tradition. Thus, we can say that Nietzsche is committed to morality in the broader sense (some prefer to use the term “ethics” to refer to broader interpretations, and reserve the term “morality” to refer to narrower, objectionable interpretations), but is opposed to various narrow interpretations of it. Since I believe that Nietzsche’s view provides the resources to retain many (re-interpreted and revised) features of morality as we understand it, I will typically use the term “morality” rather than ethics.

This dissertation has two inter-related aims. The first, and central aim, is to present and explain Nietzsche's genealogy of the concepts of guilt and punishment as a means to clarifying his view. In the process I will address various textual issues and puzzles in the text and will offer my interpretation of how they should be read and will defend those interpretations from alternate accounts. The second aim is to show how an analysis of Nietzsche's account of guilt and punishment provides the resources to critique and reject those concepts. To this end I apply Nietzsche's insights to contemporary accounts of guilt and punishment and argue that they are vulnerable to Nietzsche's criticism.

Eliminating Punishment

On the face of it Nietzsche's goal of eliminating guilt and punishment seems extravagant and implausible. The project of eliminating either guilt or punishment seems at odds with many of our own (and indeed, many of Nietzsche's own) commitments and values. One thing at stake then, is an interpretation of Nietzsche that both understands him as in dialogue with the ethical tradition itself, rather than in stark opposition to it, and that renders his philosophy internally consistent.

Consider first the idea of eliminating punishment. Nietzsche's goal seems more plausible if we think of punishment in connection to human psychology. With regard to human psychology we might take Nietzsche to be concerned with eliminating, or at least tempering our punitive impulses. Here the thought might be that we are still much too bloodthirsty in our responses to wrongdoing and crime (consider widespread support for capital punishment in our country). It seems undeniable that a great deal of unfairness

and suffering is produced through the strength of our punitive urges, and it would be a great improvement to human nature for us to have more of a desire for understanding, and more of a willingness to forgive, rather than the desire to punish.

Eliminating the concept of punishment from our understanding of the natural world seems even more plausible. Here Nietzsche could be interpreted as having the desire to remove the remaining influence of Christianity and other ascetic religions from our understanding of the natural order. On some of these views, the universe itself is understood as reflecting a moral order that punishes the wicked through death, illness and disaster, and rewards the righteous with various goods. Nietzsche would here be understood as defending a modern scientific view of the natural world by eliminating any vestiges of metaphysical notions rooted in Christianity and the ascetic ideal.

But the idea of eliminating punishment from our social institutions, seems far less plausible. On the face of it, it seems that Nietzsche is imagining either a utopian society that no longer has any need for punishment or a society that simply leaves crime unaddressed, by allowing its citizens to do as they please.³ The former seems as impossible, practically speaking, as any other utopian view and it is difficult to see how Nietzsche could believe it possible. The latter seems completely dystopian and undesirable. A society that simply decided to ignore crime would be utterly miserable

³ Those who take this view read Nietzsche's claim in GM II, in which he imagines a society with a consciousness of its own power, such that it allows itself the greatest luxury of all: letting its criminals go unpunished, as evidence that he would leave crime unaddressed. In this dissertation I explain why this view is mistaken, by pointing out the difference between leaving crime unpunished and leaving crime unaddressed. Nietzsche is for the abolition of punishment, but this does not imply that he would leave crime unaddressed.

and unjust, and it would not sustain itself for long. If this were Nietzsche's view, anyone who takes morality seriously should reject it as monstrous.

Moreover, anyone familiar with Nietzsche should recognize that this interpretation is completely at odds with Nietzsche's core commitments. One of these is a commitment to preserving civilization and culture. Throughout his work, though particularly in the *Genealogy*, Nietzsche is quite concerned to point out the looming threat to civilization that comes from the breakdown of belief in the ascetic ideal and the turn to nihilism.

The "death of God" amounts to the destruction of the ideal that served as the foundation for Western morality. Without that foundation, and a justification for a shared sense of morality, no civilization can hope to survive for long. Arguably the central concern of both *Beyond Good and Evil* and the *Genealogy* is to attract and inspire "philosophers of the future" who will be able to create new values capable of sustaining civilization and culture. But deciding to ignore crime, by leaving it unpunished, seems like a rejection of the rule of law, and seems to be just as much a threat to social stability as the loss of God and the ascetic ideal. Thus, given his concern for civilization, it is difficult to imagine how someone as smart as Nietzsche could think that leaving crime unaddressed (i.e. unpunished) would be conducive to preserving and promoting a civilization with anything like the kind of rich culture that he admires.

Eliminating Guilt

Consider next the idea of eliminating guilt.⁴ This goal seems plausible when we consider that guilt can frequently become irrational, neurotic, and debilitating. Here Nietzsche could be understood as offering a diagnosis of the modern human condition as having an over-active guilt reflex. Guilt has become so central to our psychologies that we subject ourselves (and others) to it excessively. We hold on to guilt, to the detriment of our own psychological health, and we impose guilt on others to their detriment. Additionally, given Nietzsche's well-known criticisms of Christianity (and ascetic religions generally) we might take him to be interested in eliminating the residual influence of Christianity and the forms of guilt it encourages: guilt over our physical natures, guilt over our sexual natures, guilt over our aggressive drives, etc.

But, if we understand guilt as (at least in part) the response one has in recognizing one's own wrongdoing, the idea of eliminating it from our psychologies altogether seems problematic. The capacity to feel guilt seems essential to our capacity to stand in relations of mutual recognition with others. To feel guilt over wrongdoing is to recognize the person wronged as having the standing to be treated with respect and fairness. Thus, feeling guilt amounts to recognizing that one has treated another person in ways that they do not deserve. The capacity to feel guilt also seems essential to our ability to recognize the authority of morality to make claims on our behavior. To feel guilt over wrongdoing

⁴ Here I take Nietzsche to be addressing guilt as a psychological phenomenon, rather than the legal category of guilt before the law. If he were also interested in eliminating guilt as a legal category, this would be of a piece with the view that he advocates leaving crime unpunished, and would be subject to the same criticism, it seems.

amounts to recognizing that moral demands exist, and that one has failed to live up to them.

The capacity to feel guilt seems like an indispensable condition for being a moral agent and for having anything like genuine, meaningful, interpersonal relationships. To eliminate it altogether seems to entail that we would no longer be bothered by wrongs done to others or our failure to live up to moral standards. It is hard to imagine how we could continue to be guided by moral norms, if we did not have negative responses to our own failure to do so. It is similarly hard to imagine how we could still be said to genuinely respect other people, if we did not react negatively to wronging them. Thus, if this is what Nietzsche has in mind, anyone with a commitment to morality, or genuine relationships with other people, should reject his project out of hand.

Nietzsche's Project

While these initial worries concerning Nietzsche's goal of eliminating guilt and punishment are intuitively plausible, I take them to be misguided and based on a misunderstanding of Nietzsche's view. They can be addressed in a way that both renders his project internally consistent, and his goal plausible and consistent with a robust form of morality. Nietzsche's views on guilt and punishment are complex, subtle, and inter-related. To do justice to his view will require examining both guilt and punishment in some detail.

To complicate matters further, Nietzsche claims that both the feeling of guilt and the concept of punishment originated in the context of primitive economic transactions and

the relationship between individual creditors and debtors. Guilt, Nietzsche claims, has its origins in the debtor's consciousness of material indebtedness, and punishment originated as means for an insolvent debtor to provide the creditor with repayment for a debt. Economic thinking is responsible not only for the origins of guilt and punishment, but also for their subsequent development. To understand the genealogies of guilt and punishment, from their origins to their contemporary forms, requires understanding the nature and influence of economics on the human community, and the influence of economic thinking on human psychology.

Thesis

This dissertation will draw and defend several conclusions about Nietzsche's views. Through an examination of the creditor-debtor relationship and debt consciousness, I will argue that Nietzsche provides a plausible account of early human communities and their development into the creditor-debtor relationship and trading cultures. Further, it is in this context that the conditions resulting in the feeling of "guilt" (as interpreted by the ascetic priest) are created. This occurs when the economic thinking that is central to the creditor-debtor relationship is employed to interpret "ethical"⁵ relationships and primitive (or proto-) religious beliefs and practices.

⁵ I scare-quote "ethical" here because the earliest human communities were regulated by the morality of custom. The morality of custom was the means by which the members of early human communities learned to obey the rules and customs of their community. Obedience to those rules was enforced by the dominant members of the community through aggression and the threat of violence. In its earliest stages especially, obedience to customs was an instinctive response that involved no regard for the rules *qua* rules. Members of the community felt a primitive sense of "obligation" to follow the rules, simply because they had learned to instinctively connect the thought of violating the rules with the anger and retaliation of the dominant members. This involved no distinction between "right" or "wrong" that would characterize robust ethical relationships.

Taken alone, the economic interpretation of ethical relationships (the relationship between persons living in a community) produces several questionable results. In particular, the original concept of a material debt translates onto the actions owed by an individual to the community (obedience to rules and customs) in such a way that there is no other means of addressing violations of the rules *as such*, except through the “punishment” and suffering of the rule-breaker.⁶ However, I argue that the debt model can be revised to avoid certain problems and provide a satisfactory way of understanding ethical relationships. The issue is to get past the economics of suffering and identify what is “owed” in moral terms. I defend the claim that this is regard, or respect, for moral standing, and that through the unrecognized and continuing influence of the economic model, both guilt and punishment, in their contemporary forms, are antithetical to the aims of morality.

However, a deeper and more intractable problem is produced when economic thinking is applied to early religious customs which are also understood as a source of “obligation”. The creditor-debtor relationship is interpreted onto the community’s perceived relationship to its founding ancestral spirits (and in later communities, the gods, or God) and the community understands itself to owe a debt to the ancestors, which is repaid through obedience to the customs they established and sacrifice. The economic model is a poor fit for this relationship. I argue that it is here that several unfortunate

⁶ I scare-quote “punishment” here because in early human communities punishment is not yet part of a moral interpretation of the community. “Punishment” in this context did not involve any judgments of culpability, intentions, motives, freedom of the will, negligence, etc. In this context the suffering of the rule-breaker is only seen as the means by which the debt to the community is repaid.

developments occur: The nature of debt is made increasingly obscure, “ethical” behavior is made absolutely necessary, but also doubted as sufficient to please the ancestors (or gods), religious obligations become ethical obligations, and eventually through this, the aspiration to live up to some ideal is turned into an ethical issue. These developments create the conditions in which human beings feel pervasive doubt over their inadequacies combined with anxiety over punishment, which is the underlying phenomenon that will be interpreted as guilt.

As we move into the specific set of circumstances that produce the bad conscience and the feeling of guilt I defend several claims. First, that the internalization of the instincts and compulsory obedience to new rules and customs produces the “animal bad conscience,” but is not sufficient to produce the bad conscience (consciousness of guilt). The bad conscience is a product of resentment, and these conditions are not sufficient to produce resentment. I argue that, while the slaves were forced to internalize aggression, they also had an instinct to obedience that was long-engrained in them through the morality of custom. Their aggressive drives were turned into a will to power over the expression of other drives, which allowed them to satisfy aggression through mastering themselves and becoming obedient to new customs. Thus, they adapted to their new conditions (of society and peace) without any resentment.

I argue that resentment is produced when the slaves adapt to the point that they adopt the nobles’ gods and inherit their concepts of “good and bad.” This burdens the slaves with an unpaid debt to the noble gods and the sense that they cannot provide

repayment, because they are judged “bad” according to noble ideals. This produces suffering through anxiety over punishment for the unpaid debt and through the feeling of personal inadequacy to meet their obligations. This frustrates the slaves’ will to power and produces resentment over suffering. The ascetic priest interprets this feeling as the bad conscience: a feeling of guilt before God, and punishment for sin. This interpretation redirects the slaves’ resentment towards their own animal natures as the source of their sin and suffering, and produces the dynamic of guilt. Guilt is the pervasive sense of culpable personal inadequacy that warrants, and motivates, self-punishment over one’s animal nature. This is the general account.

I argue that Nietzsche leaves us with a gap between a general, pre-historical analysis of the conditions that produce the bad conscience, and the historical events that connect it to Christianity. This connection is found in the history of post-Exilic Judaism. The exiled Jewish people suffer from the internalization of aggression through the loss of the Temple and the sacrificial system. The remaining priestly-nobles reinvent Judaism as a religion of ethical and ritual purity demanded by a holy God. The people adopt this view of God and the priestly-noble distinction between “pure and impure” (good and bad), which produces suffering and resentment when they judge themselves “impure” and thus unable to fulfill the conditions of the covenant with God. The priests interpret this as guilt before God. The slave revolt in morality begins with the Jews, when they reject the previous history of Israel and the warrior-noble values that characterized that period. The slave revolt culminates under Roman occupation, when the radically egalitarian message of

Jesus rejects the noble morality of both warriors and priests. This culminates in the doctrine of the atonement, where God takes on the punishment for sin, leaving human beings with the guilt, while also reinforcing the connection between guilt and punishment, thereby encouraging ongoing self-punishment for guilt. This is the specific account.

This specific historical account illuminates the contrast between the Judeo-Christian culture that produces pervasive guilt, and ancient Greek culture, which did not. The Greeks kept the bad conscience at bay through several means. Their gods were representations of idealized human (and animal) nature (both good and bad) and were responsible for motivating evil actions in human beings, and so functioned to “deify” the “animal in man,” rather than condemn it as the source of evil. The Greeks also internalized aggression through sacrifice, and since they did not condemn cruelty as such, the person sacrificing could see the cruelty required to do so as living up to the ideals of these noble Greek gods. Thus, internalized aggression through sacrifice and the means to self-affirmation allowed the Greeks to avoid falling into the bad conscience.

The contrast between these two cultures supports two further arguments. The first is that the conscience is sufficient to provide the kind of moral psychology that is a desirable part of a robust ethical life. The Greeks were clearly sovereign individuals who were capable of feeling pain over their own wrongdoing. They lacked the bad conscience, so the conscience is sufficient for a non-pervasive form of “guilt.” The second is that the feeling of pervasive guilt amounts to a corruption of sovereign individuality. In order to

accept the purity standards of the priests and desire to regulate their own behavior in light of those norms, the slaves of GM II had to be sovereign individuals. But through the ascetic ideal, they adopt an impossible standard that turns their sovereign individuality towards an escalating project of self-negation and self-punishment.

Thus, according to my account, Nietzsche is only concerned with the feeling of pervasive guilt as produced through the Judeo-Christian tradition. Counter to some interpreters, this does not imply that Nietzsche's account of guilt is made less relevant to persons or cultures who have moved towards atheism. I argue that what Nietzsche calls the "moralization" of guilt and duty is the identification of the animal instincts as the source of human wrongdoing. This implies that one is "guilty" when motivated by one's animal nature, and that one has a duty to "transcend" or "deny" the animal nature in order to become a genuinely ethical person. The idea of "animal nature" can be translated into more abstract ideas such as "selfishness," or "egoism," "irrational behavior," etc. and survives in secular ethical theories that emphasize altruism.

I argue that this also implies that Nietzsche's view is not opposed to what some interpreters call "locally reactive guilt," i.e. feeling pain over an instance of one's own wrongdoing. The conscience, as found in Nietzsche's "sovereign individual," enables self-regulating behavior in light of internalized norms, through an internalized connection between ethical failures and punishment. The conscience, I argue, uses painful feelings to motivate fulfilling one's obligations, and also produces pain over failure to do so. This is sufficient to preserve a sense of guilt over moral wrongdoing that is only problematic in

connection to pervasive guilt, where any ethical failure is taken to indict the person's self-worth in its entirety. Thus, Nietzsche's claim that guilt should be eliminated from our psychologies, does not entail lack of concern over wrongs done to others and a diminished form of ethical life.

My account of Nietzsche's rejection of punishment has the same outcome. Rejecting punishment does not entail the problematic conclusion that Nietzsche would have us leave crime unaddressed. Despite the fact that punishment has been interpreted as having many different expected outcomes or purposes, it has only had one means throughout, which is to inflict suffering.

The concept of punishment originates in the creditor-debtor relationship where an injury received (the unpaid debt) is taken to justify intentionally inflicted suffering in return. The parties to the creditor-debtor relationship simply accept the fact that human beings enjoy causing suffering, and may enjoy the right to cause suffering, as unquestionable. Punishment is a justification for human cruelty. This remains essential to the concept of punishment in all of its forms, and through all of its expected outcomes, however attenuated, or hidden from view. This means that the practice of punishment is in fact immoral.

But punishment is only an interpretation of already existing practices. As long as there have been human communities, there have been negative responses to those who break the rules of the community. I argue that Nietzsche distinguishes between sanctions, which are a community's way of addressing crime, and punishment, which is an

interpretation of those sanctions. Sanctions must exist to address crime, but those sanctions need not be interpreted as punitive. Sanctions can be interpreted as having the aim of protecting the community, by isolating a criminal threat from the rest of the population, and this need not have the direct aim of causing suffering, which is the aim of punishment. Nietzsche's view, I argue, goes even further than this to include the aim of rehabilitating and restoring the criminal to society, where possible.

Nietzsche's views on justice provide further evidence that the practice of punishment is immoral. Nietzsche identifies two threats to community stability: crime and resentment. Resentment over crime or injury is perhaps more of a threat to the community, because it falsifies the nature of the criminal as a means to granting the injured party the pretext to fully vent their pent-up aggression and anger. Thus, resentment leads to excesses of violent retaliation that tend to escalate. To preserve the community, those who administer justice must seek to control the expression of resentment by reserving the authority to punish to themselves. This means that legal punishment is simply a means to compromise with the demands of resentment, and is, in some sense, institutionalized revenge, however practical the results.

I argue that punishment undermines moral concerns with both the criminal and the victim. Inasmuch as the criminal lacks a conscience, punishment reinforces the belief that his methods are not wrong *per se*, but are reserved for those in power, and denied to those who are not. Similarly, inasmuch as punishment is practiced with a good conscience and satisfies the victims and authorities, the criminal sees his own motivations as no worse,

and perhaps more justifiable. Lastly, if intentionally inflicted suffering is wrong, the criminal's moral standing is ignored, in favor of (allegedly) recognizing the moral standing of the victim.

I argue further, that punishment undermines the aims of moral protest and resentment. Resentment over a wrong endured is protest over the wrongdoer's lack of recognition, and regard for one's moral standing. The aim of resentment is to bring about recognition of one's moral standing, and other appropriate responses. Punishment only anesthetizes resentment and anger, through violent affect. Thus, punishment undermines moral aims in several ways. By anesthetizing resentment, punishment provides false satisfaction and distracts the injured party from a legitimate claim to moral recognition. Punishment directs the criminal's attention to his own suffering, and thus away from recognizing the claims of those who were injured or wronged. Moreover, through the idea that suffering is sufficient to repay the crime, the criminal has no incentive to take the moral standing of the victim into account, or to take reparations seriously.

Nietzsche's genealogy reveals that the underlying phenomena that provides the basis for guilt is a feeling of relational vulnerability, that may concern other individuals, the community, or one's own relationship to ideals of human existence (religious, moral, or otherwise). The influence of economic thinking and religion distorted matters by eventually connecting relationships that may involve negative reactions from others with failures to live up to an ideal. The resulting feeling of pervasive anxiety, worthlessness and defenselessness is the characteristic feeling of guilt that concerns Nietzsche. I argue

that this moralized feeling of guilt has been the standard for understanding feelings of non-moral guilt and has led to various distortions that are problematic. I argue against the defense of “survivor guilt” as an instance of this distortion. I propose that understanding the underlying feeling as relational vulnerability is more helpful and can provide satisfying answers to the phenomenon of non-moral guilt, while at the same time recognizing that the feeling of “guilt” is often a response to the unjustified “ressentiment” of others. This recognition would, I hope, at least remove the burden of self-condemnation from the person who feels relational vulnerability in those instances.

Chapter Outline

Chapter One presents and traces out Nietzsche’s account of the development of the creditor-debtor relationship in early human communities into three distinct forms: the relationship between individual buyers and sellers, the relationship between a community and its members, and the relationship between a community and its ancestors. I present and examine the nature and development of debt consciousness in the context of these relationships and explain their importance for the later development of the bad conscience and guilt. The focus is on how primitive economic relationships influenced the way that human beings came to think about debt, obligation, and punishment. Of particular importance is the way in which economic thinking affected the religious views and practices of these communities and provided the underlying conditions for the development of guilt.

Chapter Two presents and explains Nietzsche's account of the development of the bad conscience and the feeling of pervasive guilt. The origin of the bad conscience is presented as due to the involuntary transition of human beings into the first political "state," which results in the "internalization" of the aggressive drives and the "animal bad conscience." The suffering that results from the animal bad conscience provides the initial underlying phenomenon that will eventually be interpreted as guilt, but the consciousness of guilt does not occur until resentment becomes a problem. Resentment is presented as the product of the animal bad conscience and inescapable debt consciousness, which occurs when the slaves inherit the gods and values of the nobility. This produces an impotent sense of anxiety over punishment and a sense of worthlessness, which is interpreted by the ascetic priest as guilt. The development of guilt culminates with the "temporary solution" of Christianity and the doctrine of the atonement, which only serves to exacerbate feelings of guilt and the impulse to self-punishment.

Chapter Three addresses the historical gap in Nietzsche's narrative, which occurs between his initial presentation of the foundation of the state by "some pack of blonde beasts of prey," and the culmination of guilt in Christianity. Here I present a brief history of post-Exilic Judaism and argue that it fills Nietzsche's historical gap precisely by identifying the Jewish priesthood as the nobility whose values and interpretation of God are accepted by the "slaves" of GM II. This chapter then presents the ancient Greeks as an example of a culture that avoided the bad conscience through a different use of

religion, and identifies the practice of sacrifice as the means by which their aggressive instincts were successfully internalized, and cruelty was affirmed as living up to noble ideals. Comparison of Jewish and Greek religious cultures provides an analysis of the way in which sovereign individuality was corrupted through the ascetic ideal and the bad conscience.

Chapter Four presents and explains Nietzsche's explicit comments on punishment, as well as his views on justice. Punishment is presented as an illustration of Nietzsche's view of interpretation and his distinction between the origins of a thing or practice and its meaning. This distinction shows punishment to be a particular interpretation of a community's response to threats, whether external or internal. Re-examination of the creditor-debtor relationship, reveals the concept of punishment as essentially the justification of intentionally inflicted suffering towards some end, which end is itself a matter of further interpretation. As such, punishment is based on the human desire to cause suffering. Nietzsche's claim that punishment does not awaken guilt in the criminal provides an example of his critique of punishment, which is explained and used to suggest further critique. Nietzsche's view of justice is presented and reveals that the practice of punishment is a compromise between the interests of justice in a community which struggle to control the resentment of those injured by crime. Lastly, Nietzsche's distinction between sanctions for crime and the interpretation of sanctions as punishment is used to defend the claim that punishment can be eliminated without leaving crime unaddressed by the community.

Chapter Five presents a case study that applies Nietzsche's thoughts on punishment to a contemporary account. In "Restitution and Revenge," David Hershenov rejects other theories of punishment for failing to provide justification for the "hard treatment" that is part of the practice of incarceration. Moreover, Hershenov complains that few theories of punishment pay adequate attention to the victims of crime. To address these problems, Hershenov defends a debt model of punishment connected with an unorthodox appeal to revenge. He argues that punishment provides the victims with the satisfaction of the criminal's suffering, which assuages their resentment, and restores moral equilibrium. I offer a critique of Hershenov (inspired by Nietzsche) and employ counter-examples to show that the appeal to revenge is immoral and does not justify punishment. I argue that the problem of justifying incarceration is only a problem when it assumes "hard treatment" of the criminal as necessary. Accepting Nietzsche's distinction between sanctions and punishment resolves this problem and justifies incarceration. Moreover, I argue that Nietzsche's view of justice gives us a reason to reject theories of punishment that appeal to resentment and revenge. I also argue that the debt model can be revised to satisfy moral aims, by eliminating the "currency" of suffering, and replacing it with what is actually owed: recognition of moral standing.

Chapter Six presents a case study in the application of Nietzsche's views of guilt to a contemporary account. In "Don't Worry, Feel Guilty," David Velleman offers a general theory of guilt that he applies to cases of non-moral guilt, including, especially, "survivor guilt." Velleman defends the claim that survivor guilt is a rational feeling, since it counts

as a case of “normative vulnerability,” in which the person who feels guilt feels that they have violated some normative standard and are unable to justify this violation in terms others need accept. I argue against Velleman’s defense of survivor guilt and provide counter-examples to show that his account fails on its own terms. There are no “normative standards” for surviving disasters, where one’s actions are not an issue, so “normative vulnerability” is not rational. The basic flaw in Velleman’s account is that his general account of guilt is based on a model of moral guilt, from which he attempts to explain other forms of “guilt” in those terms. His idea of “normative vulnerability” masks a deeper underlying phenomenon in human relationships, revealed through Nietzsche’s genealogy, to be “relational vulnerability” or “relational rupture,” which is the recognition that one is a source of pain to another (however innocently, or even justifiably) and thus may stand vulnerable to their negative reactions.

Chapter One: Debt Consciousness and the Creditor-Debtor Relationship

“Whence has this age-old, deeply-rooted, perhaps now no longer eradicable idea taken its power — the idea of an equivalence between injury and pain? I have already given it away: in the contractual relationship between *creditor* and *debtor*...”⁷

Introduction

As even the casual reader of GM II is aware, Nietzsche locates the origins of the feelings of moral guilt and obligation in human awareness of material indebtedness (debt consciousness). Debt consciousness in its original and earliest context is a non-moral feeling. It carries with it no suggestion that owing a debt is wrong, that failing to repay a debt is wrong, or that the one who owes and fails to repay a debt is in any way culpable for that fact. Debt consciousness develops through the activities central to primitive “economic” practices involving trade, exchange, purchase, sale, and commerce. These activities take place in the context of the creditor-debtor relationship.

Nietzsche presents the creditor-debtor relationship in three forms: between individual buyers and sellers, between the community and its members, and between the founding ancestors of a community and its currently living members. The purpose of the following section is to present the creditor-debtor relationship in these three forms, to address issues concerning the plausibility of Nietzsche’s account, and to make explicit the impact that each form of the creditor-debtor relationship has had on the development of debt consciousness. In particular, I will provide an interpretation of the text that makes sense of Nietzsche’s claim that the creditor-debtor relationship is “the oldest and most primitive

⁷ GM II:4

relationship between persons,” and that it pre-dates even the beginnings of any “societal associations or organizational forms.” I will argue that we can make sense of Nietzsche’s claims, if we apply his distinction between practice and interpretation and recognize that the economic developments that take place between individuals are subsequently used to interpret already existing things and practices in terms of debt and repayment. Thus, on my account, one central purpose of GM II is to show the effects of economic thinking on human psychology, practices, and institutions.

Looking ahead, I will argue in the next chapter that the bad conscience and the feeling of guilt is produced through the effects of internalized aggression in its connection to the individual’s sense of indebtedness to the gods (or God). The relationship to God is interpreted in a way that redirects the individual’s resentment and desire to retaliate against (or punish) the cause of its suffering, back towards the individual, which provides satisfaction to the aggressive instincts and the will-to-power.

The feeling of guilt is an overwhelming feeling of anxiety over punishment and a consuming sense of worthlessness that is interpreted as a state of punishment for sin, which thereby gives the individual the pretext to engage in self-punishment for sin. Thus, it is important to understand the dynamics of debt consciousness, how it develops, its connection to punishment, and in particular, the effects produces by its misapplication to the community’s (and individual’s) relationship to the gods (or God). Going forward I will first present the creditor-debtor relationship between individual buyers and sellers and will address several fundamental issues that arise in that context, before presenting

the creditor-debtor relationship between the community and individual, and between the community and its ancestral spirits, or gods.

1.1 The Creditor-Debtor Relationship: Individual to Individual

The first form of the creditor-debtor relationship presented in GM II is between individuals, and it is here that debt consciousness first develops. According to Nietzsche, the relationship between “buyer and seller, creditor and debtor” is the oldest and most primitive relationship between persons. “No degree of civilization however low has yet been discovered in which something of this relationship was not already noticeable.” (GM II:8) This creditor-debtor relationship is presented as a contractual relationship between buyers and sellers who engage in “making prices, gauging values, thinking out equivalents, exchanging.” Since Nietzsche claims that “the central moral concept ‘guilt’ had its origins in the very material concept ‘debt,’ (GM II:4) these activities clearly concern transactions of material goods. As long as each individual has the agreed upon “currency” at hand, there are no complications. But, buyer and seller stand in the relationship of creditor to debtor, when one party (the creditor) agrees to provide some good at hand on the basis of a promise made by the other party (the debtor) to provide future repayment. The debtor’s awareness of this outstanding material debt is debt consciousness.

But, to be debt-conscious presumes that the debtor remains aware of the debt to be repaid, and this requires that the debtor possesses a particular capacity for memory, which is not to be presumed of the earliest human beings. Nietzsche presents the creditor-debtor

relationship as pre-moral (the parties to the agreement are incapable of making genuine promises and are not motivated by moral considerations), and also presents early human beings as “forgetful animals” (GM II:1) later describing them as “partly dull, partly scattered momentary understanding...forgetfulness in the flesh.” (GM II:3) So early human beings lacked the kind of active memory (a “true memory of the will”) that would allow them to remember a debt owed.

Moreover, Nietzsche claims that this forgetfulness is due to an active and “positive faculty of suppression” (which I will refer to as “forgetfulness”) that is responsible for “screening” most of our experiences such that they enter consciousness as little as possible. So, it seems that this faculty would prevent the fact of debt from remaining present to consciousness for very long, unless there is some way of “disconnecting” the faculty of suppression for cases where something must be remembered. So, what “disconnects” this faculty and motivated the debtor to remember the debt?

The answer Nietzsche provides is: pain. “One burns something in so that it remains in one’s memory: only what does not cease to give pain remains in one’s memory.”(GM II: 3) The “mnemonics of pain” are central to the creditor-debtor relationship. Nietzsche continues:

Whenever man considered it necessary to make a memory for himself it was never done without blood, torment, sacrifice; the most gruesome sacrifices and pledges (to which sacrifices of the firstborn belong), the most repulsive mutilations (castrations, for example), the cruelest ritual forms of all religious cults...all of this has its origin in that instinct that intuited in pain the most powerful aid of mnemonics. (GM II:3)

The agreement between creditor and debtor is a case where it is “precisely a matter of making a memory for the one who promises” (the debtor). For the debtor, it is a matter of impressing “repayment on his conscience as a duty, as an obligation,” and instilling in the creditor “trust in his promise of repayment...a guarantee for the seriousness and the sacredness of his promise.” This is done through the contract which stipulates various forms of collateral in those cases where the debtor cannot repay the debt:

The debtor – by virtue of a contract – pledges to the creditor in the case of non-payment something else that he “possesses,” over which he still has power, for example his body or his wife or his freedom or even his life (or, under certain religious conditions, even his blessedness, the salvation of his soul, finally even his peace in the grave)....Above all, however, the creditor could subject the body of the debtor to all manner of ignominy and torture, for example cutting as much from it as appeared commensurate to the magnitude of the debt:— and everywhere and early on there were exact assessments of value developed from this viewpoint— some going horribly into the smallest detail — legally established assessments of the individual limbs and areas on the body. (GM II:5)

Here we see the mnemonics of pain employed by making “punishment” (not a moral concept at this stage) a central feature of the creditor-debtor relationship. In these early stages of human development, prior to the development of a moral conscience, the creditor had no reason to trust the “promise” of another forgetful animal. But an agreement to exchange present goods in return for future compensation cannot take place without some assurance of repayment. Something further is required to keep the debt present to the debtor’s consciousness, that will remind the debtor to take the debt seriously, and will oppose forgetfulness by taking the debt out of competition with any of the distracting events and experiences that will occur between making the promise and repaying the debt. The threat of suffering is what provides all of this. But, while the

desire to avoid suffering explains the debtor's motivation to repay the debt, and provides the creditor with some reason to think that the debtor is motivated to remember and repay the debt, this does not fully capture the role that suffering plays for the creditor. Note that Nietzsche lists punishment among the other items of collateral that the debtor may forfeit if the debt is not repaid. It is not obvious how punishment could function as collateral. It seems that a prudent creditor, worried about losing material goods, would only agree to transactions in which the debtor had the means to provide material compensation in some way. Punishment does nothing to provide material compensation. So why does Nietzsche emphasize that "above all" the creditor can subject the debtor to punishment?

The answer is that punishment provides the creditor with compensation for the unpaid debt by providing him with the pleasure of making the debtor suffer. The basic explanation for this is summarized in GM II:6, where Nietzsche responds to the question "to what extent can suffering be a compensation for "debts"?", with the answer:

To the extent that making suffer felt good, and in the highest degree; to the extent that the injured one exchanged for what was lost, including the displeasure over the loss, an extraordinary counter-pleasure: making-suffer...

Nietzsche recognizes that the logic of this form of compensation sounds foreign to his contemporary readers. His more detailed explanation of it is found in GM II:5, where he claims:

That in place of an advantage that directly makes good for the injury (hence in place of a compensation in money, land, possession of any kind) the creditor is granted a certain *feeling of satisfaction* as repayment and compensation, — the feeling of satisfaction that comes from being permitted to vent his power without a second thought on one who is powerless, the carnal delight of "doing evil for the pleasure of doing it" (original in French), the enjoyment of doing violence: which

enjoyment is valued all the higher the lower and baser the creditor's standing in the social order and can easily appear to him as a most delectable morsel, indeed as a foretaste of a higher status. Through his "punishment" of the debtor the creditor participates in the *right of lords*: finally he, too, for once attains the elevating feeling of being permitted to hold a being in contempt and maltreat it as something "beneath himself"— or at least, if the actual power of punishment, the execution of punishment has already passed over into the hands of the "authorities," of *seeing* it held in contempt and maltreated. The compensation thus consists in a directive and right to cruelty.

We see several things going on in this passage. The first, and most obvious, is that punishment repays the debt by providing the creditor with a feeling of satisfaction. Like any businessperson, the creditor enjoys making a profit. But the creditor also enjoys the pleasure of cruelty. This, it seems, is a basic feature of (at least early) human nature. "Seeing suffer feels good, making-suffer even more so." This, Nietzsche claims, while a hard proposition, is an old, central, powerful, and human-all-too-human one. Thus, the displeasure over the loss of material goods, is compensated for by providing the creditor with a satisfying compensatory affect (an "extraordinary counter-pleasure") — the pleasure of doing violence, and making another person suffer.

The next thing to notice is that Nietzsche places scare-quotes around "punishment." This serves to indicate that the term is not to be understood in contemporary, moral terms. Addressing the "English Psychologists" of GM II (those, like Paul Rée, who also attempt to offer genealogies of morality), Nietzsche wonders if they have ever considered the fact that "punishment as retribution developed completely apart from any presupposition concerning freedom or lack of freedom of the will?" Or that it is only after a long period of development and a "high level of humanization" that human beings are able to make

even primitive distinctions between “intentional,” “negligent,” “accidental,” “accountable,” and their opposites and to apply these concepts when “measuring out punishment.” Or that the seemingly natural, instinctively obvious thought that “the criminal has earned his punishment *because* he could have acted otherwise,” is a late, and very sophisticated development that cannot be attributed to early human beings. If the English Psychologists did not lack historical sense, they would have recognized that:

“Throughout the greatest part of human history punishment was definitely *not* imposed *because* one held the evil-doer responsible for his deed, that is, *not* under the presupposition that only the guilty one is to be punished: — rather, as parents today punish their children, from anger over an injury suffered, which is vented on the agent of the injury...” (GM II:4)

Thus, “punishment,” in the context of the early creditor-debtor relationship is non-moral. There is no assumption that the debtor intentionally broke the terms of the contract, or entered into the contract with the intention to deceive the creditor, or was culpably negligent by underestimating his ability to provide repayment, or culpable for taking on a risky venture without doing due diligence. Suffering punishment is consistent with the debtor simply having bad luck with regard to his ability to repay the debt. Punishment in this context has nothing to do with addressing a moral wrong. It only serves to provide the creditor with compensation for his loss, through the pleasure of making the debtor suffer.

Note, too, Nietzsche’s comments on the relation of social standing and status to the pleasure the punishing creditor feels. The lower the creditor stands in the social order, the greater satisfaction he feels in punishing the debtor. This is connected to having

permission to do violence and hold another person in contempt — as beneath one. This serves to give the creditor a feeling of elevation, the “foretaste of a higher status.”

The connection between punishment and social standing indicates several things. The first is that the community in which the creditor-debtor relationship exists has some form of authority structure and hierarchy. It is possible to stand higher or lower in the social order, and violence within the community is the prerogative of those who stand higher in that order (it is a “right of lords”).⁸ Thus, it seems that violence within the community is not generally acceptable and must be “justified” in some sense (however crudely) in order to receive permission.

The second thing to notice is that the creditor-debtor relationship has developed into a means to raising one’s social standing. The creditor granted the opportunity to punish is not merely engaging in an activity usually reserved for those who rule and enforce customs (though he is doing that). The creditor is given a “foretaste of a higher status.” The creditor was able to either produce or acquire, sufficient surplus goods to be able to grant loans to other members of the community. The fact that the creditor is allowed to punish defaulting debtors suggests that the community itself recognizes the value of the

⁸The language used here — “contempt,” “right of lords,” “authorities” — and the very mention of social standing and feelings of “elevation,” appears suggestive of the nobles that are discussed in GM I. It is the nobles who conquer and establish a hierarchical political structure between those who rule and those who are slaves. It is the nobles who feel “elevated” above the slaves and hold them in “contempt” as base and common. Thus, these terms lead scholars like Daniel Conway, in *Nietzsche’s On the Genealogy of Morals* to think that the creditor-debtor relationship between individuals occurs after the founding of the political state. If this is the case (as we will see) then it implies that all forms of the creditor-debtor relationship developed within the context of the political state. (Conway, 2008) I think Conway is mistaken and will later show that it is both inconsistent with the text, and that the language used here can be explained as consistent with pre-political communities.

creditor's production or trade skills as a means to producing wealth (in some form) and the injury that is caused by a defaulted loan. Thus, contractual relationships are recognized as practices that must be protected by the community's customs, and wealth has become a source of power and status, that can be gained or lost. Being allowed to punish effectively communicates to the creditor, and the rest of the community, that wealth is important to the community, and those who have wealth have greater standing.

This suggests that the community has undergone significant development in terms of economics. At least some trade between community members is no longer a straightforward exchange of goods. Some members have surplus goods that others lack and desire to obtain. These debtors are at least capable of imagining a future in which they are able to produce or obtain the goods required to repay the debt. This suggests a community that has grown to the point that makes possible, and probably requires, a greater degree of division of labor, and so requires the kind of evaluative, calculative thinking central to the creditor-debtor relationship in order to negotiate more complicated transactions of goods.

Finally, it should be noted that the practice of punishing debtors is regulated by what I will call the "equivalence principle", which develops within the creditor-debtor relationship. When Nietzsche compares punishment of the debtor out of anger over an injury received with parents punishing their children out of anger, he adds:

Anger held within bounds, however, and modified through the idea that every injury has its equivalent in something and can really be paid off, even if only through the pain of its agent. (GM II:4)

This principle is the product of calculative and evaluative thinking that involves “making prices, gauging values, thinking out equivalents, exchanging,” which, Nietzsche claims “preoccupied man’s very first thinking to such an extent that it is in a certain sense thinking itself.” The pervasiveness of economic thinking dominates early human beings to the point that:

The eye was simply set to this perspective: and with that clumsy consistency characteristic of earlier humanity’s thinking – which has difficulty moving but then continues relentlessly in the same direction. (GM II:8)

This thinking produces what Nietzsche calls “the oldest and most naive moral canon of justice,” the generalization that “every thing (*jeder ding*) has its price; everything (*alles*) can be paid off.” (GM II:8) So, the influence of the creditor-debtor relationship functions to institutionalize material losses as injuries that warrant retaliation.

The picture Nietzsche presents is of an early community in which retaliation over injuries received is accepted (whether this is conducted by the injured party, or by those who rule the community, seems irrelevant in this context). The development of the creditor-debtor relationship alters this in two important ways. The first is by recognizing economic losses as injuries. But, perhaps more importantly, through the ability to make comparative evaluations regarding the equivalences between material goods, the creditor-debtor relationship produces the equivalence principle, which imposes customs, and limits on retaliation. In Nietzsche’s terms, we see a new form imposed on older practices. Retaliation over an injury has been interpreted as punishment (still in a non-moral sense), and because the opportunity to do violence is itself a valued commodity that can serve as

compensation for a debt, punishment is seen as “deserved” (also in a non-moral sense; “apt,” or “due” might be the more precise term). It is one way to pay off a debt that must be paid off, and – in the minds of earlier humanity, it really does pay off the debt.

1.2 Problems with the Creditor-Debtor Relationship

There are several apparent problems with Nietzsche’s account of the creditor-debtor relationship that need to be resolved before we proceed. Resolving these issues first will make the relationship between the three forms of the creditor-debtor relationship clearer and more plausible. This, in turn, will make the development of debt consciousness, through the influence of economic thinking, clearer, and will highlight Nietzsche’s methodological distinction between practice and interpretation.

1.2.1 Memory & Promising

Nietzsche’s goal in GM II is to provide a genealogical account of the feeling of moral obligation and moral guilt. These feelings are to be traced back to their origins in completely natural events and material concepts. GM II presents a contrast between early human beings, who lack these feelings, and the example of the sovereign individual. The sovereign individual represents the end product of the “morality of custom,” which has succeeded in creating an active memory in the individual, and gives him the permission to make promises. (GM II:2) This contrast is meant to direct our attention to the way that the morality of custom produced changes in early human nature that lead to the sovereign individual as the end result.

The sovereign individual is “permitted to promise,” because he is able to make the fact that he has made a promise a source of ongoing motivation. The sovereign individual feels obligated to keep his promises, because they are promises (and would, presumably, feel guilt over failure to do so). If asked to identify the source of these feelings, the ability to make genuine promises, the knowledge of the extraordinary privilege of responsibility, and consciousness of a rare freedom, the sovereign individual would say: my conscience.

But the sovereign individual’s ability to make genuine promises is based on many other skills and capacities. In particular, it presumes the possession of an active memory (a “true memory of the will”), that can work in opposition to the human being’s active faculty of suppression (forgetfulness). (GM II:1) The active memory and other capacities that are required for genuine promising, and constitute the sovereign individual’s conscience, are the product of a long history of development, achieved through the harshness of the “morality of custom.” The sovereign individual’s “conscience” is nothing more than this active memory and other capacities that have sunk to an unconscious level and have become a dominant instinct. Thus, the sovereign individual is a very late product in the history of human development.

In contrast with the sovereign individual, early human beings were forgetful animals who lacked an active memory. Since an active memory is a pre-requisite for a conscience, they also lacked a conscience and any feelings of moral obligation or guilt. Nietzsche’s project is to begin with the capacities they *did have* and to show how these capacities

were transformed in various ways and are eventually interpreted as moral guilt and obligation.

This story begins with the “oldest and most primitive relationships between persons there is”: the creditor-debtor relationship. (GM II:8) The individual buyers and sellers in this relationship are early human beings, so they are forgetful animals who lack consciences, do not feel moral obligation, and are incapable of making genuine promises.

But note that when Nietzsche presents the collateral pledged in the creditor-debtor relationship he claims that this is “in order to instill trust in his *promise* of repayment, to provide a guarantee for the seriousness and the sacredness of his *promise*, to impress repayment on his *conscience* as a *duty*, as an *obligation*...” (GM II:5, my emphases) If the concepts mentioned here are genuinely moral concepts, then Nietzsche, in less than one full page, has gone from berating the “English Psychologists” for basing their genealogies of morality on moral concepts (GM II:4) and tacitly assuming precisely what they are trying to explain, to doing the same himself.

To see that this is not the case, we need to make a few distinctions. The first is between two senses of “memory” in the text. This distinction is found in GM II:1 where Nietzsche claims that the kind of will that can oppose the active faculty of suppression and “disconnect forgetfulness for certain cases – namely...where a promise is to be made” is not:

Simply a passive no-longer-being-able-to-get-rid-of the impression once it has been inscribed, not simply indigestion from a once-pledged word over which one cannot regain control, but rather an active no-longer-wanting-to-get-rid-of, a

willing on and on of something one has once willed, a true memory of the will.
(GM II:1)

Thus, being “permitted to promise” like a sovereign individual, requires that one has an active memory, a “true memory of the will.” The early human beings Nietzsche presents in the creditor-debtor relationship do not have this capacity. But the fact that they do not possess a “true memory of the will” does not preclude the possibility that they have the capacity to recall events from the past and make promises, in the sense of entering into contracts.

Recall that Nietzsche’s goal is to trace the development of moral guilt and obligation from their original non-moral sources in the creditor-debtor relationship. Nietzsche’s language of “creating” or “making” a memory can mislead us into thinking that he is giving an account of memory as such. But if we pay attention to the idea of a “true memory of the will” we see that this is not the case. The issue is not the origins of the faculty of memory; the issue is the origins of moral motivation. Nietzsche is not trying to explain how human beings developed the convention of promising, or how they remember facts about the past. Nietzsche is trying to explain how promising became a source of motivation – how it came to be felt as an obligation.⁹ And, as we saw earlier, the motivation for early human beings was provided by the prospect of punishment.

⁹ Bernard Reginster summarizes it this way: “Conscience is a particular kind of memory, which Nietzsche calls ‘the will’s memory’. This is more than the memory of the fact that I once willed to do something as commentators suppose. It is also the perpetuation of the motivation to do as I willed – ‘a desire to keep on desiring what has been, on some occasion, desired (GM II:1). Conscience, therefore, is a motivational structure, which moves me, for example, to follow through on my promises.” (Reginster, 2018, p. 4)

The connection between promising and punishment, once internalized and made instinctive, explains the feeling of obligation. This same connection explains the feeling of guilt in cases where the obligation is not fulfilled. As human beings are conditioned through punishment to take promising seriously, they adapt to this new demand. They learn, and become more routinely able to keep their promises. As keeping promises becomes habitual, the active faculty of suppression causes any conscious and explicit connection between promising and punishment to be forgotten. The connection is incorporated into the person's unconscious instincts and promise-keeping becomes an instinctive activity. But promise-keeping remains motivated by the now instinctual feeling that breaking a promise is connected with suffering. Thus, the anticipation of suffering and the desire to avoid it is the feeling of obligation, and the anticipation of suffering that one cannot avoid because of a broken obligation is the feeling of guilt.

1.2.2 The Conscience

Paying attention to distinctions in the text has provided us with a way to defend Nietzsche's appeal to memory in his account of the creditor-debtor relationship from the charge of "question-begging." But Nietzsche also mentions the conscience in this account. He claims that provisions for punishment in the contract are a means for the debtor to "impress repayment *on his conscience* as a *duty*, as an *obligation*." [my emphases] As stated earlier, the conscience is a late development found in the sovereign individual and should not be presumed in early human beings.

To defend Nietzsche from the charge of “question-begging” here, we need to pay close attention to his presentation of the mnemonics of pain. The mnemonics of pain is first presented in GM II:3, and it is clear that the mnemonics of pain are employed through punishment in the creditor-debtor relationship. (GM II:5) Careful examination of these passages will reveal that there are two different senses of “conscience,” and “obligation” employed.

Read in the order they appear, GM II:3 is a discussion of the mnemonics of pain as a means of motivating obedience to the community’s basic social rules, while GM II:5 is a discussion of the mnemonics of pain as a means to motivate promise-keeping in the context of the creditor-debtor relationship. So, there is at least *prima facie* evidence to consider the possibility that these are distinct stages of human development. Examination of the text will show that this is Nietzsche’s view.

GM II:5 is entirely about the creditor-debtor relationship and primarily about the way that punishment provides the injured creditor with compensation. In this context, punishment is clearly attached to a “contract,” which is entered into voluntarily by both parties through making an agreement (the creditor to advance goods, the debtor to repay that advance). Moreover, given Nietzsche’s claim that guilt has its origins in the “very material concept “debt,” (GM II:4) the transaction taking place here involves material goods; it is a purely economic transaction.

GM II: 3 has a completely different focus. GM II:2 ends by revealing that the sovereign individual's dominant instinct is his conscience. The topic of the conscience resumes with:

His conscience? ... One can guess in advance that the concept 'conscience' which we encounter here in its highest, almost disconcerting, form, already has behind it a long history and metamorphosis. (GM II:3)

The passage then proceeds to describe the oldest means to this metamorphosis: the mnemonics of pain, which, as noted above, serves to burn something into one's memory through its association with pain.

If one is familiar with the entirety of GM II, it can be tempting to read this passage as simply providing the general theory behind the practice of punishment in the creditor-debtor relationship. But closer examination calls this into question. Whereas the creditor-debtor relationship involves voluntary agreements and promises between individuals, GM II:3 is solely concerned with obedience to social rules. After claiming that rituals involving blood, torment, sacrifice, and mutilations were employed whenever it was necessary for early humans to make memories for themselves, Nietzsche connects this with remembering "a few primitive requirements of social co-existence." He later says that, with the images provided by the aforementioned rituals and processes, "one finally retained in memory five, six 'I will nots,' in connection with which one has given one's promise in order to live within the advantages of society." There is no mention of creditors, debtors, contracts, or economic transactions.

One might read the text as first presenting the sovereign individual's conscience and permission to promise (GM II:2) through an active memory ("true memory of the will") (GM II:1), before explaining its genesis through the mnemonics of pain. (GM II:3) On this reading, the text then reveals the context in which the mnemonics of pain first produced a memory of the will and the conscience: the creditor debtor-relationship. (GM II:5)

But Nietzsche's presentation of the creditor-debtor relationship claims that "precisely here there are promises made" (GM II:5), and that punishment was employed in part to "impress repayment on his [the debtor's] conscience as a duty." This reading threatens Nietzsche with inconsistency. It has Nietzsche appealing to the conscience and promise-making as central to his explanation for the development of the conscience and the permission to make promises.

The text suggests an alternative reading. Nietzsche claims that genuine promising requires an active memory to counter active forgetfulness. (GM II:1) The result is that "a world of new strange things, circumstances, even acts of the will may be placed without reservation between the original "I want," "I will do," and the actual discharge of the will, its act, without the long chain of the will breaking." But then he points out how much an active memory and the capacity for genuine promising presupposes. These presuppositions are that:

Man must first have learned to separate the necessary from the accidental occurrence, to think causally, to see and anticipate what is distant as if it were present, to fix with certainty what is end, what is means thereto, in general to be able to reckon, to calculate, — for this, man himself must first of all have become

calculable, regular, necessary, in his own image of himself as well, in order to be able to vouch for himself as future, as one who promises does!

Nietzsche goes on to claim that:

...the task of breeding an animal that is permitted to promise includes, as condition and preparation, the more specific task of first making man to a certain degree necessary, uniform, like among like, regular, and accordingly predictable. (GM II:2)

This work was accomplished by the morality of custom, which is the “true work of man on himself for the longest part of the duration of the human race, his entire prehistoric work.” (GM II:2) Nietzsche is claiming that the morality of custom first made human beings calculable, etc. both in their natures and in their own self-understanding, and this was what led to their capacity to think in calculative terms. Thus, making humans calculable occurred before the creditor-debtor relationship where they first employed calculative thinking.

We have seen that the parties involved in creditor-debtor relationships are not yet capable of making genuine promises. This is why the threat of punishment is included in the contract. But the individuals in the creditor-debtor relationship *are* capable of thinking in calculative terms. This implies that they have already been made calculable, etc. and this was accomplished through the morality of custom. Thus, the text suggests several stages of development: One in which humans are made calculable, etc. through the morality of custom, another in which humans learn to think in calculative terms (still in the context of the morality of custom), and later, through the introduction of other

factors, a stage in which this behavior is made instinctive, and produces humans who are sovereign individuals and permitted to promise.

If this is correct, it allows us to explain Nietzsche's reference to the conscience in GM II:5 as consistent with his claims about the sovereign individual. GM II:3 addresses the way that the mnemonics of pain was employed to "inscribe" the memory of a few social rules on human consciousness. This is the earliest work of the morality of custom, which makes humans conform to the customs of the community, and thereby makes them regular, predictable, etc.

The work of the morality of custom is largely accomplished through the community's dominant individuals whose instinct is to command. Enforcing customs, through threat of violence, produces a psychological connection between violating customs and pain. This provides the community's members with the motivation to avoid pain through obedience to customs. This motivation is the earliest feeling of (non-moral) obligation and the primitive memory that sustains this form of motivation can be thought of as the earliest form of the "conscience." We find support for this in BGE 199, where Nietzsche claims:

Inasmuch as ever since there have been human beings there have also been human herds...and always very many who obey compared with the very small number of those who command...hitherto nothing has been practiced and cultivated among men better or longer than obedience, it is fair to suppose that as a rule a need for it is by now innate as a kind of formal conscience which commands: 'thou shalt unconditionally do this, unconditionally not do that', in short 'thou shalt'. This need seeks to be satisfied and to fill out its form with a content.

The idea of a "formal conscience," which is an innate need to obey, allows us to explain the mention of a "conscience" in GM II:5. The events of GM II:3 are much earlier than

the beginnings of the creditor-debtor relationship presented in GM II:5. The first work of the morality of custom is to produce animals that are obedient to the primitive requirements of social co-existence. Without obedience to these basic rules, cooperation is impossible, and the creditor-debtor relationship could never develop.

The individuals in the creditor-debtor relationship are capable of calculative thinking. Thus, the existence of the creditor-debtor relationship in GM II:5 presumes that the long work of making humans calculable, etc. and obedient to customs has already been largely successful. This means that the humans in the creditor-debtor relationship already have an instinct to obey primitive customs. Thus, they have a “formal conscience,” which can be described as a memory of basic rules and customs. But, the formal conscience seeks to be filled out with content. In the creditor-debtor relationship, the mnemonics of pain (the practice of punishment) is a means of inscribing new customs pertaining to economic transactions and contracts on the debtor’s formal conscience. It is a means of making a “memory of the will,” which can be described as memory of one’s promises, and reenforcing the feeling that obedience to the new custom of promise-keeping is “obligatory.”

1.2.3 The Oldest Relationship Between Persons

Having addressed the worry that Nietzsche’s account of the creditor-debtor relationship is guilty of “begging the question” with regard to his use of the concepts of “memory” and “the conscience,” I now turn to other features of Nietzsche’s account that are not issues of inconsistency, but of straightforward plausibility. In order of

presentation, these are Nietzsche's claims that: the creditor-debtor relationship is the "oldest and most primitive relationship between persons," that the creditor-debtor relationship is "older than even the beginnings of any societal associations and organizational forms," and his implied claim that human beings are naturally cruel.

The claim that the creditor-debtor relationship is the "oldest and most primitive relationship between persons that there is" (GM II:8) is baffling. One would think that humans first related to each other as parents, children, mates, friends, or even enemies, rather than as buyers and sellers. So, if Nietzsche is not making an exceedingly implausible claim here, there must be something else he has in mind.

The term "persons" can be read as synonymous with "individual human beings," but can also be read as marking the difference between "human being" as a biological category and "personhood" as a normative category. The latter marks a distinct stage of development, or achievement in human nature – a status that is not automatically attached to being human, and which some *homo sapiens* never attained. I take Nietzsche's use of the word "person" to be marking this distinction. Thus, on this read, the individuals who made up the earliest human communities were *not* persons in the *normative* sense of the word and only became persons *through* the creditor-debtor relationship.

So, what counts as "personhood" for Nietzsche and how does the creditor-debtor relationship produce it? The answer is that personhood is constituted by the capacity to think in terms of values and reasons – the ability to engage in reasoning, reflection, abstraction, and imaginative thinking, and the motivation to guide one's actions thereby.

This, it seems, is what sets human beings apart from other animals.¹⁰ And it is through the economic activities found in the creditor-debtor relationship that humans first come to think in this way.

We find textual support for this in GM II:4, where Nietzsche refers to reason, mastery over the affects, and reflection as the “prerogatives and showpieces of man.” This theme is resumed in GM II:8 where Nietzsche claims that economic thinking is perhaps the beginning of “human pride, man’s feeling of pre-eminence with respect to other creatures.” This is in no way a full account of Nietzsche’s doctrine of personhood, but it is sufficient for present purposes.

Nietzsche also explicitly claims that economic thinking “preoccupied man’s *very first thinking*.” [my emphasis] While earlier human beings were not absent thought, inasmuch as they were sentient beings responsive to features of their environment, (including the expectation of others), it is through the creditor-debtor relationship that they first begin to exhibit thinking that can be thought of as primitive instrumental reasoning, valuing, and reasons-responsiveness. In order to enter into contracts and exchange goods, the participants had to be able to assess the value of those goods, negotiate a price, and imagine (or project) a future where the terms of the contract are fulfilled.

Nietzsche also claims that it was in the creditor-debtor relationship that “for the first time a person *measured himself* by another person.” This suggests that it was here that

¹⁰A full account of Nietzsche’s view of personhood would require an extended discussion of his drive-psychological doctrine of the soul. I am not claiming that the capacity to value and think are sufficient conditions for personhood for Nietzsche. But it is sufficient for present purposes that both of these passages indicate that thinking and valuing are features that appear to set human beings apart from other animals.

human beings began to think in terms of comparative self-worth. The development of the creditor-debtor relationship provides a plausible explanation for this and marks an important turning point in human culture. It is the moment when material wealth becomes a means of power and standing within the community.

As noted above, human beings were conditioned by the morality of custom to obey rules, long before the existence of the creditor-debtor relationship. These human beings were herd animals. (BGE 199) Harsh conditions made belonging to the herd a requirement for survival. Remaining with the herd required sensitivity to the herd's expectations, and in particular, obedience to those who commanded and enforced customs. Behaving differently than others was perceived as dangerous, and so was not advantageous to the individual's survival. Thus, in this earlier time, there was both an incentive to not stand out from the others – to only compare oneself in negative terms (“don't be different, don't stand out”), and also very little likelihood for most members to have the power and status of those who dominated and ruled over the community.

The creditor-debtor relationship changes all of this.

Recall that the existence of the creditor-debtor relationship implies that the community has developed significantly, such that some members have surplus goods that they are willing to part with, on the promise of future repayment. This implies that the community has grown in size, stability, security, and strength, such that it can sustain activities that aim at more than mere subsistence and survival. This further implies more sophisticated divisions of labor and an increase in specialization.

Specialization, along with more time to spend on matters that do not concern immediate survival, creates the possibility of innovation that allows for production or acquisition of more goods. Those that can produce those goods are now seen as different, in a way that is advantageous to the community.

Previously, the community relied on those who had the strength and aggressiveness to defend the community from external threats and enforce customs and cooperation internally. These dominant individuals provided the benefits of living in a community: safety, security, and internal stability. But when these efforts are successful – when the community is secure from external threats and internally stable through obedience to social rules – the need for violent, aggressive action becomes less directly, or obviously relevant to the everyday functioning of the community. Instead, commerce, trade, and the ability to negotiate transactions become the central activities of everyday life. And those who are successful in commerce become a new source of advantages to the community. So the development of the creditor-debtor relationship provides a new way for individual members to gain power and raise their social standing.

Engaging in negotiations, making deals, and succeeding or failing in these endeavors makes possible the beginnings of comparative evaluation and self-worth. As buyers and sellers negotiate they are forced to evaluate and put a price on their own goods, as well as other individuals' goods. They are forced to recognize that the other individual does the same. In some primitive sense, they are forced to think from the other's point of view. Those who have more desirable goods, or are more skillful negotiators, and are more

successful, will gain a sense of their own skills, and power, in comparison to others, and with that a sense of pride and superiority over others in these matters.

Thus, we see that the conditions and activities required for the creditor-debtor relationship provide a plausible explanation for the claim that it was the first place that “person stepped up against person” and “measured himself by another person.” The creditor-debtor relationship requires the individuals in it to be capable of: primitive instrumental and causal reasoning, thinking in terms of the value of goods, imagining a future that can be brought about through their own actions, and recognizing others as doing the same. Moreover, through evaluation of their own skills and success in comparison to others, these individuals can see themselves as having greater or lesser social standing, which affects their sense of pride and self-worth. All of this seems to satisfy the requirements for primitive personhood.

1.2.4 Older Than any Societal Associations or Organizational Forms

The preceding explanations of Nietzsche’s use of terms like “memory” and the “conscience”, as well as his claim that the creditor-debtor relationship is the “oldest relationship between person,” all rely, either tacitly or explicitly, on one crucial assumption: the existence of a community in which various conditions and practices produce a memory, or a memory of the will, or the formal conscience, or the creditor-debtor relationship itself. But Nietzsche makes claims that seem to undermine that assumption. He can be read as claiming that the creditor-debtor relationship between individuals developed prior to any community, and was the basis for establishing the

community. This would run counter to my account. More importantly, if this is the case, then Nietzsche presents individuals entering into contracts without any means to guarantee that the terms of repayment will be enforced, which produces a completely implausible Hobbesian social contract view of the community.

The problematic passage in question is this:

Purchase and sale, together with their psychological accessories, are older than even the beginnings of any societal associations and organizational forms: it was out of the most rudimentary form of personal legal rights that the budding feeling of exchange, contract, guilt, right, obligation, compensation first transferred itself onto the coarsest and earliest communal complexes (in their relationship to other similar complexes, together with the habit of comparing, measuring, and calculating power against power. (GM II:8)

It is one thing to claim that economics and the creditor-debtor relationship were an early, and fundamental, feature of human relationships, and had an enormous effect on human moral and psychological development. But to claim that it predates “even the beginnings of any societal associations or organizational forms” is implausible. It suggests that human beings first interacted as individuals in the creditor-debtor relationship before there were any forms of human society.

Even a brief consideration of the creditor-debtor relationship should make this implausibility obvious. The creditor-debtor relationship begins with the debtor’s “promise” to repay goods advanced by the creditor. But, since both parties, at this stage, are “forgetful animals”, who lack a “true memory of the will,” this is not a genuine promise. The creditor cannot rely on the debtor to be motivated by the mere fact of

having promised to repay. This is why the contract is backed up by the debtor's pledge of collateral, and includes the threat of punishment.

To agree to transactions of goods and to put up collateral presumes that both parties recognize the concept of property. Both parties recognize certain goods as the creditor's property, and other goods as the debtor's property to pledge as collateral. The items of collateral mentioned further support this. To pledge one's wife is, at this stage of human development, to pledge one's property. To pledge one's freedom points to the practice of slavery and other human beings as property. And it is implausible to think that either marriage practices and the practices of slavery could exist without a community to recognize and enforce norms that make property ownership possible.

The debtor may also pledge his blessedness, salvation, or peace in the grave. This implies the existence of religious beliefs and practices and it is implausible to think that these developed outside of a community. More to the point, if religious consequences are an explicit part of the contract, this implies that the creditor and debtor share similar beliefs. It seems unlikely that they would do so, unless they also shared a community.

Most importantly, the contract includes, and relies on the threat of punishment. The debtor's only motivation to provide repayment is the threat of future loss or suffering. The creditor's willingness to accept the debtor's promise is based on the assurance that he will be compensated in some way – even if it is only through the debtor's suffering. And if the contract is only between two individuals, punishment, and thus repayment is not assured. Fulfillment of the contract would either rely on the creditor's own strength to

wrest the collateral away from the debtor, or on the debtor's willingness to surrender the collateral, or submit to punishment. In the pre-moral context that Nietzsche assumes, the debtor has no reason to do either, and in fact has more reason to break the contract and keep the creditor's goods, if he has the strength to do so. For punishment to provide a genuine threat, and genuine motivation in the creditor-debtor relationship, there must be a community to enforce the terms of the contract and punish the debtor for non-repayment.

Thus, if Nietzsche claims that the creditor-debtor relationship began between individuals, before the community existed, but his account of the creditor-debtor relationship requires a community to recognize and enforce norms, and carry out punishment, then his account is internally inconsistent, and the creditor-debtor relationship is impossible as a pre-moral phenomenon.

To resolve this problem we need to pay attention to the language that Nietzsche uses to refer to different types of human social groups. He claims that the creditor-debtor relationship predates even the beginnings of any "societal associations and organizational forms" [*gesellschaftlichen Organisationformen und Verbänden*]. (GM II:8) Later in the same passage he claims that the activities and "psychological accessories" that develop in the creditor-debtor relationship were first transferred onto "the coarsest and earliest communal complexes" [*Gemeinschaft*] which are organized according to blood relationships. Later in the text, Nietzsche claims that consciousness of debt to the gods (which will be fully presented and explained in the following sections) does not come to an end "even after the decline of the community organized according to blood-

relationships.” (GM II:20) This passage follows his claims about the origins of the first political “state,” where:

...some pack of blond beasts of prey, a race of conquerors and lords, which, organized in a warlike manner and with the power to organize, unhesitatingly lays its terrible paws on a population enormously superior in numbers perhaps, but still formless, still roaming about. (GM II:17)

Thus, the comments in GM II:20 seem to clearly reference the transition from nomadic groups that were organized according to blood-relationships to the organized political state. It appears then, at least from these passages, that Nietzsche uses the terms “societal associations and organizational forms” to refer to politically structured societies, and uses the terms “communal complexes” and “community” to refer to nomadic groups organized by blood relations, which had no formal political structure.

Further support for this distinction can be provided through other pieces of text. First, in GM II:5, Nietzsche claims that the “commune” is the most primitive (and one would then assume, earliest) form of society. The use of the term “communal complexes,” combined with the fact that Nietzsche is explicitly discussing the earliest stages of human development, suggests that he might have something like the commune in mind. The fact that he claims that these communal complexes were organized by blood relations, further suggests that he is speaking of nomadic family units, or tribal groups, and it seems that in their earliest, smallest, and most primitive stage, these types of groups would have had a fairly communal structure. Survival would require that many, if not most of the group’s activities were cooperative endeavors, and goods produced cooperatively would be shared communally. (Later, as these groups grow in size and greater division of labor is

possible, those developments will put pressure on the community to develop practices of barter and trade.)

Recall, too, Nietzsche's claim that, "Inasmuch as ever since there have been human beings there have always been human herds (family groups, communities, tribes, nations, states, churches)..." (BGE 199) Given Nietzsche's admiration for the political organization of the Roman Catholic Church into a hierarchy from the most spiritually qualified to the least (in principle, if not in practice)(GS 358), it is plausible to read this list (family groups, communities, tribes, nations, states, churches) as ranked from most primitive to most politically sophisticated. If this is so, then it provides further support for the claim that Nietzsche treats the community (as well as family groups, tribes, and nations) as forms of society that predate the politically organized state. I want to suggest that these pre-political social groups fit the picture of nomadic groups in increasing degrees of size, development, and sophistication (family groups, communities, tribes, nations) that range from the earliest hunter-gatherers to larger trading cultures.

Taken together, this evidence provides a plausible solution to the current problem. Nietzsche claims that the creditor-debtor relationship began between individuals first, and existed long before the beginnings of the first politically organized societies. But these individuals were not isolated Rousseauian loners. Prior to the existence of the state, there were primitive, nomadic, communal complexes, and it was in the context of these communities that the conditions for the creditor-debtor relationship developed and that relationship originated. And, subsequent to its development, the creditor-debtor

relationship, and the economic thinking central to it, was used to interpret other existing practices and relationships within those communal complexes, such as the practice of banishment, the individual's relationship to the community, and the community's relationship to the ancestral spirits, or gods. (Keeping this instance of the distinction between practice and interpretation in mind, will make understanding the development of the creditor-debtor relationships between the community and its members, and between the ancestors and the community much clearer.)

1.2.5 Human Cruelty

A final issue must be addressed before proceeding to the other forms of the creditor-debtor relationship. As we have seen, punishment makes transactions possible in the creditor-debtor relationship by providing forgetful debtors with the motivation to keep their promises to pay back debts. But punishment also provides the creditor with compensation for unpaid debts. If no other material goods are available, or equal to the debt, the creditor can take it out of the debtor's flesh – literally. And punishment can only provide compensation if punishing others is something that the creditor enjoys and values.

The function of punishment in the creditor-debtor relationship turns on the claim that the creditor (and humans generally) enjoys the pleasure of inflicting suffering on other human beings. Since the creditor-debtor relationship develops long before the existence of the politically organized state, this cruelty cannot be explained as originating through the repression and internalization of aggressive animal drives. The foundation of the state

explains the internalization of aggression that produces self-cruelty, but, antecedent to that development, humans enjoyed inflicting suffering on others. Thus, Nietzsche must be claiming that human beings are naturally, or innately, cruel.

The claim that human beings are naturally cruel, and seek opportunities to inflict suffering on others, is in tension with the idea that the earliest communities needed to cooperate in order to survive, and likely had a commune-like social structure where goods were cooperatively gained and shared. This does not sound like the behavior of cruel animals, nor does this behavior seem possible for animals that are naturally cruel.

This raises a plausibility issue. As we have seen, Nietzsche's account of the creditor-debtor relationship is only plausible if we assume the existence of a community to sustain it. In order for a community to function well and survive its members must have some sense of reciprocity and a disposition to look out for each other's safety and survival, i.e. some primitive instinct for "altruistic" behavior. The issue is whether this is consistent with Nietzsche's claims about human cruelty.

This worry can be addressed by considering the conditions of existence for early human communities and the traits that would give these groups an evolutionary advantage, as well as Nietzsche's comments on the earliest responses to the "criminal," i.e., the individual who breaks the group's customs.

Natural selection favored communities whose members had strong instincts for aggression. These instincts, through the natural conditions of warfare, produced human cruelty. But natural selection would also favor communities whose members had a strong

sense of reciprocity toward each other, which could be described as primitive “altruistic behavior.” The instinct for aggression and cruelty is directed towards threats to the community, while instincts for reciprocity characterize the internal relations of the community.

Consider first, natural selection and aggression. As we saw, the earliest human communities were extended family units, perhaps consisting of several generations of members. When herd instinct prevails and individuals remain with the community there is a greater chance of survival for those individuals and the community itself. Within these family units it is plausible to assume the existence of stronger, dominant members who give commands and enforce the behavior of the other members, similar to the behavior we see in our closest non-human relatives. Call these the “alphas.” At the earliest stages, it is likely that the alpha was dominant because of greater strength and greater capacities for aggression and violence, simply because these attributes would make them best suited to protect the other members from external threats (enemies, predators), and better able to deal with internal rivalries and disputes. Natural selection would favor family units with stronger, more aggressive alphas who successfully protected the community and commanded obedience within.

Living in harsh conditions, these communities faced (at the very least) the need to hunt for food and protect themselves from dangerous predators, which requires aggressive behavior and violence. Natural selection would favor communities whose members (beyond the alpha) had aggressive instincts and a capacities for violence,

making them better able to acquire food, protect the community, live long enough to procreate, and pass these instincts on to their progeny.

But this only accounts for aggression, which does not yet amount to cruelty. Cruelty is not necessary for hunting, and in fact, is counter-productive. The goal of the hunt is to procure food and this is best accomplished by killing the game quickly, rather than taking time to make it suffer.

Yet, Nietzsche claims that human beings are not only cruel, but *celebrated* cruelty. In connection to the creditor-debtor relationship, he claims that making-suffer (through punishment) is a “true festival” and that “cruelty constitutes the great festival joy of earlier humanity, indeed is an ingredient mixed in with almost all of their joys.” Moreover, since this cruelty occurred in pre-moral context, it was “naive and innocent” and “something to which the conscience heartily says “yes”!” Nietzsche adds, “without cruelty, no festival: thus teaches the oldest, longest part of man’s history — and in punishment too there is so much that is festive!— ” (GM II:6)

Careful examination of these claims reveals something important. Nietzsche claims that cruelty constituted “the great festival joy of earlier humanity,” but also says that “in punishment *too* there is so much that is festive.” [my emphasis] This suggests that festival cruelty is something that precedes the practice of punishment, and that punishment is accepted by the community because it recalls, or re-creates that earlier festival. If this is the case, then punishment becomes a practice because it satisfies an already existing desire for cruelty that is connected with festival celebration.

A festival suggests a community-wide celebration. So, what events would the entire community have cause to celebrate? A successful hunt is a possibility, but as stated earlier, it is not very plausible to see cruelty as an essential feature of the hunt. However, victory over the community's enemies would be another occasion for celebration, and in the context of war, cruelty becomes quite plausible.

It is likely that early human communities frequently found themselves in competition with other communities over scarce resources. Given natural selection, it is likely that the members of these other communities were also aggressive and capable of violence. As such, they would pose a very real threat to the community's survival and be seen as enemies of the community. If these rivals are strong enough to pose an on-going threat, and scarcity eliminates the option of migrating elsewhere, war becomes inevitable. And the conditions of war are conducive to the development of cruelty.

Fighting an enemy involves facing an individual that one can recognize as intent on doing one harm. One natural, instinctive response to a direct, and intentional attempt to do one harm is anger and the desire to retaliate. But retaliation can be a purely defensive response that aims at preventing further harm to oneself. This might require violence, but need not involve cruelty.

To see how cruelty develops we need to think in Nietzsche's terms. The fear, anxiety, and anger that are part and parcel of violent conflict, or imminent threat, are strong affects. As affects they create an emotional tension in the individual who feels them, which seeks discharge. If the community has learned obedience to customs, this instinct

will prevent violence within the community, but this restriction only serves to add to the tension felt. Thus, when the community has an opportunity to discharge these affects externally, toward the enemy, their expression will be commensurately violent. The enemy will receive the full brunt of pent-up affect and retaliation will seek to make the enemy suffer as much as possible. Discharging violent affects relieves pent-up tension and produces pleasure. Thus, the retaliating party learns to associate making suffer with pleasure, and cruelty just is taking pleasure in making suffer.

The idea that cruelty begins in war finds further support in the text, where Nietzsche describes the treatment of the “criminal” within a community as:

The copy, the *mimus*, of normal behavior toward the hated, disarmed, defeated enemy, who has forfeited not only every right and protection, but also every mercy; in other words, the law of war and the victory celebration of *vae victis* [woe to the conquered] in all their ruthlessness and cruelty. (GM II:9)

Here we see that cruelty toward the enemy is the original festival and that punishment of the criminal recalls that festival cruelty. (Later we will see that, in the earliest stages of “justice,” this is “justified” by making the criminal an enemy of the whole community, so on a par with an external enemy.)

Appeal to the conditions of war provides a plausible explanation for the development of cruelty as a feature of human nature. Natural selection, and the requirements for a stable community explain why cruelty does not threaten the existence of the community.

Reciprocity can be explained in similar terms. The arrangement of the community into a primitive command-obedience structure serves the interests of all the members. The alphas are obeyed, not only because they are stronger, but because they protect the

group. The price of this protection is obedience. To disobey the alpha is to risk expulsion from the group and the protection it provides. By providing this protection, the alpha gains obedience from the other members, and through governing their relations with each other, thereby increases the effectiveness of the group and his own prospects for survival.

A primitive sense of reciprocity between members (which need only be instinctual at this stage), would provide a greater evolutionary advantage to these communities.

Members who do not learn to engage in reciprocal behavior are a threat to cohesion.

Allowing this behavior undermines the community's ability to cooperate and coordinate their activities effectively, such that its members no longer know what to expect from others. This decreases the community's ability to provide protections and benefits to its members. Alphas who are instinctively motivated to command obedience, and to expel disobedient members, preserve group cohesion. Over time, this selects for members who are largely responsive to the rules and engage in reciprocal behavior. communities that largely consist in members who engage in reciprocal behavior function more effectively and can provide protections and benefits to most of its members. These communities have an evolutionary advantage over communities whose members lack an instinct for reciprocity.

This seems to require that the community's alpha also has an instinct for reciprocity, both with regard to their own behavior and the behavior they expect from the other members. Alphas lacking a strong instinct to protect the community offer the members little protection, and limited chances of survival. But, just as importantly, alphas who do

not expect and enforce reciprocity between the other members fail to preserve the group's cohesion and ability to cooperate, which threatens group survival. Thus, natural selection favors communities whose alphas have a sense of reciprocity, expect the same from other members, and enforce norms of reciprocal behavior.

Ultimately, there is no inconsistency between the claim that human beings are naturally cruel and also naturally capable of reciprocity and "altruistic" behavior. The conditions for survival in the earliest stages of human development favored those who had aggressive instincts and capacities for violence. The context of war provided conditions that connected the expression of those instincts with other pent-up affects, and the pleasure of discharging them against the enemy. Had conditions been less harsh, this connection might not have been established, but as a contingent fact, the pleasure of discharging one's violent affects was associated with doing violence to others.

The conditions for survival also favored groups whose members felt a sense of reciprocity towards each other. Human beings were too weak, compared to many other animals, to survive alone, and only survived through the instinct to remain in groups. But a collection of individuals can only function as a community if it can cooperate and coordinate its activities well enough to provide protection and benefits to most of its members. And it can only cooperate effectively if there are a shared set of customs and expectations concerning behavior within the group. The alphas are responsible for "creating" these expectations by commanding obedience from the other members.

Communities whose alphas enforced norms of reciprocity, and whose members were more responsive to those norms, were more effective and stable, had a greater ability to provide protection and benefits to most members, and greater chances of survival. Thus, natural selection favored communities whose members were capable of both cruelty to enemies and reciprocity towards other members.

1.3 The Creditor-Debtor Relationship: Community and Individual

Having resolved the issues raised for the creditor-debtor relationship between individuals, we are now in a position to present the other forms of that relationship. In addition, we are in a position to show how the relationship between individuals, which developed first, was used to interpret other existing practices and relationships.

The second creditor-debtor relationship presented in the text is the relationship between the community and its individual members. Nietzsche introduces it by beginning with:

Always measured by the standard of an earlier time (which earlier time is, by the way, at all times present or again possible): the community, too, thus stands to its members in that important basic relationship, that of the creditor to his debtor.
(GM II:9)

The basic idea of this form of the creditor-debtor relationship is that the community provides its members with various benefits:

One lives in community, one enjoys the advantages of community (oh what advantages! We sometimes underestimate this today), one lives protected, shielded, in peace and trust, free from care with regard to certain injuries and hostilities to which the human outside, the “outlaw” is exposed... (GM II:9)

By providing protection and other advantages, the community functions as a creditor who advances goods to its members without receiving payment up front. This makes the members of the community debtors, who owe the community repayment. The form of that repayment is understood to be the member-debtor's obedience to customs.

Obedience to customs is the "price" set on the advantages the community provides.

The community protects its members from injuries and hostilities that those outside the community experience. The understanding is that, as a member, "one has pledged and obligated oneself to the community precisely in view of these injuries and hostilities." So, each member has a contract with the community, which provides protection and advantages, in return for which the member is obligated to obey customs. The individual who breaks the community's customs – the "criminal" – is a debtor who "not only fails to pay back the advantages and advances rendered him, but also even lays a hand on his creditor." The community sees this as "least of all a matter of the direct injury inflicted by the injuring party," and is more concerned with the fact that "the criminal is above all a 'breaker', one who breaks his contract and word with the whole in relation to all goods and conveniences of communal life in which he has until this point had a share." The response of the community is what one would expect from any deceived creditor. The community "will exact payment as best it can," by seeing to it that the criminal "not only forfeits all of these goods and advantages from now on, as is fair, — he is also now reminded *how much there is to these goods*." The community does so by banishing the criminal from the community, and making him an "outlaw" in the original sense of the

word: one who is given no protection under the “laws” of the community, and upon whom “every kind of hostility may vent itself on him.” (GM II:9) As noted above, “punishment” at this stage is a copy of the normal manner in which the community would treat a defeated enemy.

The structure and internal dynamics of the creditor-debtor relationship is clear in this case. There is an agreement (a contract) between the community and its members. The community, as creditor, advances goods to its members. Members, as debtors, owe repayment in the form of obedience to customs. Those who fail to provide repayment are punished, and the community’s motive in doing so is the pleasure of retaliating against an enemy. Since the criminal received various advantages and protection from the community, the criminal forfeits the same: all advantages and all protection.

But there are other aspects of the relationship to the community that are not as clear, and there are ways of reading these passages that lead to misunderstandings of Nietzsche’s view. Paying careful attention to the text will help us to avoid those misunderstandings and will also bring out subtle details that are quite important.

Take, for instance, the opening phrase “always measured by the standard of an earlier time.” What earlier time is Nietzsche speaking of, and what standard does he refer to? This is important, because it is clear from the passage that this earlier standard is the basis for construing the community and its members as standing in the creditor-debtor relationship.

One might take the “earlier time” in question to be a time before the community mentioned in the passage. Given the order of presentation, this would be the time of the events in GM II:8, where the creditor-debtor relationship between individuals is described in most detail. Here individuals begin to think in terms of “the most rudimentary form of personal legal rights” and the “budding feeling of exchange, contract, guilt, right, obligation, compensation first transferred itself onto the coarsest and earliest communal complexes.” The standard in question could be the “equivalence principle”, which Nietzsche claims is the “oldest and most naive moral canon of justice”, which is the idea that “every thing has its price; everything can be paid off.”

This can lead to the idea that, for Nietzsche, the creditor-debtor relationship between individuals develops before the community, and then the feelings, concepts, and economic thinking central to the creditor-debtor relationship are employed to *form* a community through an explicit agreement. Thus, the community *begins* with its members thinking of it as the creditor and each member as a debtor, and uses the reasoning found in the equivalence principle to justify the practice of punishing those who break their contract with the community, through breaking the basic social rules.

This appears to find further support when we recall the language Nietzsche uses in GM II:9. The individual is understood to have “pledged” himself to the community, precisely in order to receive its advantages and protection. The criminal is treated as “one who breaks his contract and word with the whole.” This language implies that the relationship between the community and the individual is a matter of an explicit

agreement – a contract entered into voluntarily, with explicit terms that spell out the advantages and expected repayment. Thus, it can appear that Nietzsche is describing a community that forms and continues to grow through making explicit contracts with its members.

But this clearly cannot be the case. We have seen that the creditor-debtor relationship between individuals required an already existing community as a necessary condition of its possibility. Thus, the community must have existed before the creditor-debtor relationship and before its members began to think in terms of creditors, debtors, contracts, and promises. We have also seen that the earliest communities were extended family units and complexes based on blood-relations. Membership in these communities would have mostly been a matter of birth into the community, which continued through the member's instinct to remain, conditional on obedience to the rules. Most of this is not voluntary: birth into a community, the current rulers, the rules imposed, are obviously not chosen, and certainly not the result of an explicit contract or promise.

So, if this is not inconsistency on Nietzsche's part, something else must be going on. The previous interpretation, while incorrect, does contain some truth. The earlier time in question *is* when the creditor-debtor relationship between individuals developed, and the standard in question *is* the equivalence principle. (GM II:8) But the relationship that these have to the "earliest and coarsest communal complexes" is a matter of *interpretation*, rather than *formation*, as the previous interpretation had it. The already existing community, its members, customs, and practices are all *interpreted as* another instance of

the creditor-debtor relationship. This is what Nietzsche means when he says that the concepts and economic thinking that developed between individual buyers and sellers was “transferred” *onto* these communities. The creditor-debtor relationship led the members to apply economic thinking to their relationship to the community, interpreting the community as creditor, and the advantages it provides as goods advanced, and themselves as debtors, whose obedience to customs is repayment of the debt.

There are subtleties in this interpretation that are important. The creditor-debtor relationship between buyer and seller begins with a contractual agreement to exchange goods or services. When the buyer contracts to receive goods and provide repayment this transforms buyer into debtor, and seller into creditor. Thus, the creditor-debtor relationship comes into existence through a voluntary agreement. While one party might feel pressure, or have less leverage in the transaction, the deal itself is still voluntary.

As we have seen, the earliest communities did not originate through an explicit agreement, or contract. And yet, the account of the creditor-debtor relationship to the community, as we have seen, includes the idea that the individual has pledged himself to the community, and that the criminal is treated as a contract-breaker. Since Nietzsche rejects the idea that *the state* began through a contract as a “flight of fancy” (GM II:17), it hardly seems likely that he endorses a social contract account of the earliest and most primitive communities.

So, if these communities did not begin through contracts and agreements, the most plausible explanation for why they begin to think in those terms is the influence of the

creditor-debtor relationship. The community already provides benefits to its members – and they would be aware of this. Obedience to customs is already expected, and most members are well aware of this. Those who cannot obey are banished from the community. The creditor-debtor relationship (“man’s very first thinking...that...is in a certain sense thinking *itself*” (GM II:8)) is applied to these existing practices. The community is interpreted as a creditor that advances goods. No creditor simply advances goods (there is no “for nothing” in this “soul-poor age” (GM II:19)) The community wants something in return. Obedience is interpreted as repayment for goods advanced. One cannot be either creditor or debtor without entering into an agreement. Therefore, the relationship between the individual and the community is interpreted as a matter of a contract that was voluntarily entered into.

This same “reasoning” is applied to reinterpret the sanctions employed by the community against customs-breakers (initially, banishment). Enforcing obedience to customs long preceded the origins of the creditor-debtor relationship. The text suggests that banishment was the original form those sanctions took. Initially, banishment was an instinctive “herd response” to members that threaten the community through breaking customs. Economic thinking changes the way that this practice is understood.

In the creditor-debtor relationship between buyer and seller, the creditor only has the right to punish if two conditions have been met: The debtor has voluntarily entered into a contract and the debtor has failed to repay the debt through material compensation. But

where the debtor has so failed he has broken the contract; punishment is due, and the creditor has a right to it.

When the individual's relationship to the community is reinterpreted as a matter of a voluntary agreement, this makes disobedience to customs a matter of a broken contract. The practice of banishment is then interpreted as a "warranted" response by the community to those who break their contracts with the whole. If the customs-breaker is a debtor who refuses to provide repayment (in the form of obedience to customs), then, since suffering is one means of repaying a debt, suffering is due, and the opportunity to cause suffering is owed to the community. The community has a right to make the customs-breaker suffer through banishment, because there is a contract between the community and its members that the criminal has broken. Banishment is then seen as having the aim of depriving the contract-breaker of the illicitly gained advantages provided by the community, and providing a reminder of those advantages by allowing the community to vent its anger on him.

This is the origin of the retributive sense of punishment that Nietzsche alludes to in GM II: 4, when he wonders whether the "English Psychologists" have ever considered the possibility that "punishment as retribution developed completely apart from any presupposition concerning freedom or lack of freedom of the will" and long before human beings were capable of making, and applying, distinctions between " 'intentional,' 'negligent,' 'accidental,' 'accountable,' and their opposites" when "measuring out punishment." Early human beings caused suffering out of anger over an

injury received, and because they enjoyed doing so. Through the creditor-debtor relationship, this primitive, instinctive retaliation is interpreted as “justified,” as the appropriate response to customs-breaking, which causes suffering as a means to settle a debt. Thus, through the economic thinking of the creditor-debtor relationship, instinctive retaliation in anger and the practice of banishment is transformed from a primitive social sanction into a “justified” pre-moral version of retributive punishment.

Thus, we can distinguish a time when the practice of banishment should not be considered punishment, because the creditor-debtor relationship did not yet exist and had not provided an interpretation of this practice. In the early stages, banishment functioned as a sanction (whether those who practiced it realized this, or not). It had the effect of removing dangerous elements from the community and preserving social stability. There is a common tendency to think that non-punitive banishment was simply replaced by punishment, which took other, less severe, forms. But the text indicates that this is not exactly correct.

Note that when Nietzsche presents the practice of banishment (“outlawing”), he describes the community as thinking that depriving the criminal of the advantages of society is *fair*. [no scare-quotes, my emphasis] Moreover, the community’s desire is to remind the criminal of “how much there is to these goods,” by leaving him open to every kind of hostility that can befall those outside the community. The aim, it seems, is to make the criminal suffer for his crime – both physically and emotionally. Thus, banishment was interpreted as having the purpose of both depriving the criminal of the

community's advantages and causing suffering. It is "fair," because it is "deserved" (in a non-moral sense), and it is "deserved," because the criminal has not properly paid his debt, through obedience to customs. This makes it a form of punishment (though this is still in a non-moral sense), because it has "justified" the direct aim of making the criminal suffer for his crime.

However, banishment is eventually replaced by other forms of punishment. Nietzsche claims that:

As its power grows, a community no longer takes the transgressions of the individual so seriously because they can no longer count as dangerous or subversive for the continued existence of the whole to the same extent as formerly... (GM II:10)

This indicates that the community has come to see banishment as too severe, and "unjustifiable" with regard to many crimes, because those crimes no longer threaten the entire community. Population growth, for instance, would make it the case that no single person's failure to contribute goods to the community would endanger it. With more people contributing goods, the criminal has less impact on overall welfare, and is more of a free-rider than a direct threat.

Here the creditor-debtor relationship has had further influence. Breaking customs was originally seen as primarily an issue of a broken contract with the whole community, because customs-breaking threatened the welfare and survival of the whole community. But in the creditor-debtor relationship between individuals there would be recognition of greater and lesser degrees of debt, such that a broken contracts had different degrees of significance in terms of losses incurred, and demanded different degrees of suffering

owed as repayment. As the community grows it applies this thinking to broken customs. It recognizes that some crimes produce lesser harms than others and only affect some individuals, rather than the entire community. Banishment makes sense when all are threatened with harm, but can no longer be seen as appropriate when it is only individuals who are harmed. Thus, the mere fact of a broken contract with the community ceases to be the main criteria by which the community's response is seen as "fair." The harms produced and losses incurred must be taken into account as the criminal's "debt," which in turn dictates the amount of suffering the criminal must endure to provide repayment.

At this stage, banishment is replaced with other forms of punishment that are judged "equivalent" to the crime in question. The community also begins to deny general anger any opportunity to vent itself on the criminal. (GM II:10) This shift in perspective is the result of the economic thinking found in the creditor-debtor relationship that has produced the generalization that "every thing has its price; everything can be paid off," (GM II:8) such that every form of injury – including injuries to the community – has an equivalent in something, even if it is only through the suffering of the agent of injury.

Note that this shift from banishment to other forms punishment is the first instance of humans making comparative evaluations about the relative worth of actions. Initially, all crimes threatened the community and so banishment appeared appropriate. But economic thinking has led the community to see differences in value between actions. Crimes have varying degrees of significance. Once the community is understood as a creditor providing advantages, and obedience to customs is understood as repayment, then any

form of debt is capable of being repaid, and different forms of punishment replace banishment as the means of doing so. Thus, debt consciousness develops to include the idea that crimes can, and should be paid off – just like material debts – and punishment is the appropriate means for repayment.¹¹

1.3.1 Impact on Debt Consciousness

Punishment as practiced in the creditor-debtor relationship between buyer and seller turns on the idea that the debtor's suffering repays the debt to the creditor. Thus, adopting punishment indicates that the community now views individual crimes as capable of being paid off through suffering. In principle (and at least for less serious crimes), it seems possible for the criminal to no longer be viewed as an enemy of the community but as a member that can be restored, once the crime is paid off. Indeed, Nietzsche claims that those who are responsible for administering "justice" in the community begin to protect the criminal from the anger of the injured party, and the general anger of the

¹¹ Here we should note a point that is central to the critique of punishment. Banishment originally functioned as a sanction that removed dangerous elements from the community and preserved social stability. So, while the earliest communities banished out of anger and cruelty, one way that banishment might have been interpreted (had things gone differently) is as a sanction that is solely meant to protect the community by removing a threat. The community might have then progressed to slowly recognize that different actions have differences in significance in terms of their threat to the community and its members, and might have moderated its response to different threats accordingly, by replacing banishment with other ways of neutralizing a threat: containment through incarceration, education, treatment, etc.

Instead, economic thinking interpreted banishment as providing repayment through suffering, transforming it into punishment, rather than a sanction. The purpose of punishment is not to remove a threat (however much it achieves that result in the process), but to cause suffering; and suffering is interpreted as repaying a debt. But punishment only provides repayment because it satisfies human cruelty. Thus, the economic interpretation serves to establish the idea that crime (or wrongdoing generally) deserves suffering and establishes a connection between debt/guilt and suffering that are both reinforced over millennia. This only obscures the fact that punishment is nothing more than a justification for cruelty.

community, and limit responses to crime to established equivalents. (GM II:10) This, at least, seems to be a positive result of the shift to punishment.

But the fact that crimes have been interpreted as sources of debt to the community expands the nature of debt for the individual. The contract between buyer and seller only involves one way of going wrong: failure to repay the debt in the terms specified by the contract. But in relationship to the community, all of an individual's actions count as either conforming to custom or breaking it. Thus, there are many more ways to fail to repay one's debt to the community.

The nature of the debt is also altered. The debt between buyer and seller is both finite and voluntary. It exists when individuals come to an agreement concerning a transaction of goods. The debt can be decisively paid off, whether through the agreed upon currency, collateral, or punishment. And once the debt is repaid, by whatever means, the debtor is relieved of the burden of debt consciousness. Moreover, if one is successful financially, one may never need to go into debt again. Material debt is escapable.

The debt owed to the community is finite, but less "voluntary" in various ways. Though the community may understand its members as having entered into a contract with it, through which they are made debtors, it seems that many, if not most, members are born into debt, simply by being born into the community and immediately receiving its benefits.¹²

¹² Perhaps the community practices some form of "maturity ritual," through which the individual formally acknowledges membership and a debt to the community. But it seems that this amounts to publicly acknowledging a relationship that already exists without an explicit agreement having been made beforehand. And if these rituals pre-dated the creditor-debtor relationship, this would be another instance of reinterpreting an already existing practice in its terms.

The relationship to the community seems less voluntary than the relationship between buyer and seller in other ways. In at least some cases, the individual interested in “taking out a loan” might have options with regard to possible creditors, including the option of declining to enter into a contract at all. And, if the debtor does decide to enter into a contract, it will typically be a matter of negotiation. Sellers may set prices on their goods, but sellers need buyers as much as buyers need their goods. There will be at least some situations where the buyer has an advantage and can negotiate better terms.

The community, however, has a monopoly on the advantages it provides. It is the only creditor to its members and its terms are non-negotiable. The goods offered by the community are set, as is the price of those goods: obedience to the community’s customs. The stakes for declining to come to an “agreement” with the community are much greater. To decline membership is to decline *all* of the advantages of the community (as opposed to the advantages of one business deal with an individual creditor) and would amount to self-banishment.

In addition, the debt owed by the individual is a function of the power of the community and the advantages it is able to provide. As the community grows in power and is able to provide greater advantages, the individual’s debt grows as well. The average individual will have very little influence over this growth. (This factor becomes especially fraught in its connection to the ancestors and gods.)

Finally, the benefits provided by the community are on-going, so repayment of the debt is also on-going. As long as one remains a member of the community one can never

be completely “out from under” this debt. This is an important (and it seems overlooked) feature of Nietzsche’s account of debt consciousness. The individual who takes himself to be indebted to the community always carries (however unconsciously) the weight of debt consciousness, and with it, the threat of punishment. Thus, it is here, in the creditor-debtor relationship to the community that debt consciousness begins to be an ever-present feature of human psychology.

1.4 The Creditor-Debtor Relationship: Ancestors to Community

The final form of the creditor-debtor relationship is the one that is perceived to exist between the ancestors of the community (in particular the founders) and its currently living members. Nietzsche describes it this way:

Within the original clan associations – we are speaking of primeval times – the living generation always acknowledges a juridical obligation to the earlier generation, and particularly to the earliest one, which founded the clan... (GM II: 19)

This “juridical relationship” is the relationship between creditor and debtor. Nietzsche states this explicitly, when he notes that “the civil-law relationship of the debtor to his creditor...was once again...interpreted onto...the relationship of those presently living to their ancestors.” The living members of the community are indebted to their ancestors, because it was through their efforts, sacrifices, and achievements, that the clan exists.

But, the early community’s relationship to their ancestors is not one of fond memories and warm gratitude towards their benefactors. The “founders” of the community were the already mentioned alpha-figures, who aggressively and violently defended the community from external threats and enforced obedience and cooperation within the

community. The community believes that they continue to exist as powerful spirits. Thus, the founders were intimidating figures who were feared by the community while they were alive, and continued to be feared in their spiritual existence.

The influence of the creditor-debtor relationship makes the next step quite natural for early human beings. Once the community is seen as a source of benefits to its members, those who founded the community are seen as a source of benefit as well. And, as powerful spirits, the founding ancestors “do not cease to use their strength to bestow on the clan new benefits and advancements.” By providing advantages, the ancestors are creditors to the entire community. The idea that they might do so out of purely altruistic feelings is completely incomprehensible to the people of these “soul poor ages.” There is no “for nothing.” The ancestors expect repayment for the advantages they provide, thus they are feared just as any creditor is feared: As a potential source of punishment, should the community fail to repay its debt.

It is likely that the original alphas demanded a share of any goods produced or acquired by other members, in return for the protection they provide. In their continued existence as spirits they expect what they’ve always expected: obedience and a share of the community’s goods. Thus, the “fear of the progenitor” that Nietzsche mentions in GM II:19, motivates the community to continue following customs and providing the ancestors offerings and sacrifices, in order to avoid the ancestor’s angry reprisal.

Belief in the ancestors, then, adds another source of debt. Through establishing the community the ancestors have already provided a great, and on-going, source of benefits

to the community's members, in the form of protection, peace, trust and freedom from various injuries and hostilities. The community as a whole enjoys these advantages, so there is a collective debt to the ancestors.

As with the community, these advantages create a debt that is not static – it continues to accrue, because the ancestors are present and provide new advances. Any growth in strength by the community, along with any ability to provide greater advantages to its members, is taken as evidence that the ancestors are using their power to benefit the community, so the debt to the ancestors grows as well.

Repayment of the debt comes in the form of “sacrifices.” Initially, Nietzsche claims, these sacrifices may have been innocuous. They were the literal things that the ancestors might have expected, were they still alive: Nourishment (presumably in the form of food offerings), holding festivals in their honor, building shrines to commemorate them (likely to literally house them as well), offering tribute (perhaps a portion of the spoils of war, or the hunt), and above all – obedience to the customs they established, which are treated as their statutes and commands.

Over time, however, with the growth of the community and a greater sense of debt, the community begins to suspect that these innocuous sacrifices are not pleasing to the ancestors, because they are not proportionate to the magnitude of the debt. The community worries that, as angry creditors who have not been properly repaid, the ancestors might withhold their aid, or even retaliate against the community. They begin to offer greater sacrifices to the ancestors. As this cycle continues, and suspicion continues

to grow, the community develops a felt need to redeem itself by providing an “enormous counter-payment,” which Nietzsche suggests might involve sacrifice of the first-born, but at the very least, requires blood – human blood especially.

Nietzsche suggests that it is perhaps this perceived relationship to the ancestors that is the origin of belief in the gods. This idea gains plausibility when we consider the distinction between practice and interpretation. In this case, the relationship to the ancestors and the various practices involved, are interpreted, and reinterpreted, over time to produce new conceptions of the ancestors, and cultic practices that reflect the community’s own self-understanding. As a community grows, its success is taken as evidence of the ancestors’ continuing aid. The magnitude of that success – when the community is victorious, independent, honored, and feared – is evidence of the magnitude of the ancestors’ power. This increases the community’s fear of the ancestors. The growing distance in time between the original founders and the current community erases any actual memory of those figures, and they slowly become legends and myths. The more the ancestors recede into the distant past, and the more successful the community is, the more it is impossible for the community to believe that their current state is the result of merely human endeavor. Eventually:

...through the imagination of growing fear, the progenitors of the most powerful clans must have grown into enormous proportions and have been pushed back into the darkness of a divine uncanniness and unimagibility; — in the end the progenitor is necessarily transfigured into a god. (GM II:19)

Thus, the conception of the ancestors is interpreted in light of the community’s own situation and self-understanding. These conceptions reflect the community’s political

structure as well. According to Nietzsche, universal empires tend to believe in universal deities, and despotism tends to prepare the way for some form of monotheism. The concept of God reaches its culmination with the “Maximal God,” which Nietzsche specifically equates with the Christian God. And with the concept of the Maximal God, there is a maximum of debt.

1.4.1 Implications for Debt Consciousness

When Nietzsche presents the equivalence principle he comments that:

The eye was simply set to this perspective: and with that clumsy consistency characteristic of earlier humanity’s thinking — which has difficulty moving but then continues relentlessly in the same direction — one arrived straightaway at the grand generalization ‘every thing has its price; everything can be paid off’ — at the oldest and most naive moral canon of justice... (GM II:8)

This hints at the idea that the economic thinking found in the creditor-debtor relationship (already questionable through the idea that suffering is a form of currency) may not be applicable to other forms of relationships. Nietzsche claims that it is a human tendency to reduce the unfamiliar to familiar concepts and categories. (BGE 192 & 230) The creditor-debtor relationship and economic thinking is an early example of this tendency. We see its ruthless (mis)application especially clearly in the relationship to the ancestors.

Some features of the creditor-debtor relationship to the ancestors are familiar from the other forms. As with the other forms, goods are advanced and repayment is expected. As with the relationship to the community, the relationship to the ancestors is not voluntary. The community cannot choose its ancestors, or the customs they established. The debt is also non-negotiable. As long as the ancestors provide benefits, the community owes

obedience to customs and sacrifice. The debt is also on-going and constant. The community cannot get out from under this debt while it is still successful.

But the creditor-debtor relationship of the community to its ancestors results in developments that are unique and strangely at odds with the other forms of the creditor-debtor relationship.

One significant difference between the relationship to the ancestors and the other forms of the creditor-debtor relationship is that the ancestors are not present to their debtors in the way that the community or individual creditor is present. Unlike an angry seller who has not been repaid, or a community that is upset with an individual's behavior, the ancestors are not available to make their demands and desires known, and it cannot be verified whether they are satisfied with the payment offered or not.

This leaves the form and amount of repayment to the fearful imagination of the community. Fear of the ancestors introduces the suspicion that they have not been satisfied, and with every success of the community, this suspicion grows and debt consciousness increases. Thus, the feeling of debt to the ancestors introduces anxiety over debt and doubt over one's ability to provide satisfactory repayment. The idea that debt continues to grow produces an increased sense that one is incapable of discharging that debt, which increases anxiety over punishment.¹³

¹³ I am in agreement with Conway on this point. He notes that the connection between the gods and the success of the community leaves the community with no way of anticipating its future obligations and involves no limit to the magnitude of debt that may accrue, and that this is a significant development for understanding the shift from debt consciousness to guilt consciousness. (Conway, 2008, p. 87)

Through anxiety, the repayment owed to the ancestors becomes less straightforward and more mysterious, more subject to doubt. The fact that the ancestors are spiritual beings contributes to the doubt. Initially they are understood as content with the kinds of goods they enjoyed while alive. But as they are seen to grow in power, they become more uncanny and inscrutable, as do their desires. What exactly does a spirit want? What will appease a god? The community begins to doubt that its offerings are adequate. They ask “Does one ever give them enough?...the suspicion remains and grows...” (GM II:19)

Thus, we see that the debt to the ancestors is not only an on-going pressure within debt consciousness, it is also escalates in intensity. Each success of the community is evidence of the ancestor’s power and increases the debt. This increases the suspicion that they have not been duly repaid. This suspicion leads to greater sacrifices (blood, castration, the first-born), which appears to create an escalating spiral of debt consciousness, doubt, and greater, crueler sacrifices as the community attempts to rid itself of the pressure of unpaid debt, and anxiety over punishment. Thus, debt consciousness approaches an extreme, where great and terrible sacrifices will be necessary to redeem the community from debt.

To make matters worse, the relationship to the ancestors, for all its anxiety over punishment to the community, does not interpret the community’s misfortunes as punishment that could be understood as paying off the debt. Nietzsche claims that the fear of the progenitor’s power increases with the success of the community but,

...every step toward the atrophying of the clan, all miserable chance occurrences, all signs of degeneration, of approaching dissolution always diminish the fear of

the spirit of the founder and give an ever more reduced notion of his shrewdness, his foresightedness, and his presence as power. (GM II:19)

This effectively alters the role that punishment plays in the relationship to the ancestors.

In the relationship between individuals, or to the community, punishment is understood to provide definitive repayment for debt. But in the relationship to the ancestors it ceases to have this function. If specific misfortunes are not interpreted as punishment from the ancestors, but rather, as a sign of their weakness, then the community lacks one means that other debtors have to be relieved of debt consciousness. As long as the community exists and is successful, it is in debt to the ancestors, and the only sign that the debt might not exist, would come at the cost of losing the ancestors, and the end of the community itself. Thus, the community is trapped in a situation in which it must will to be in debt, if it is to will its continued existence and success. Fear of punishment over debt remains, but by eliminating specific misfortunes as instances of punishment, it remains an unfocused anxiety that can never be relieved.

This would indicate that the community does not see their ancestors as interested in relieving them of debt consciousness, which leaves the community with no way to address its debt except through its own efforts. One way they attempt this is by following the rules and customs commanded by the ancestors. But anxiety over debt produces a tension here. On one hand, it serves to dramatically increase the sense of seriousness with which customs must be observed, but on the other, it reduces obedience to customs to a merely necessary condition for repaying the ancestors.

Since the ancestors are understood to be the source of the community's continued success, collective anxiety over providing repayment creates pressure on each individual to follow the customs they established. Failure to follow customs becomes a serious matter, since disobedience invites the displeasure of the ancestors and threatens the entire community. This leads to an increasing emphasis on the seriousness of following customs. (Nietzsche goes so far as to claim that the community feels a collective responsibility over the individual's violation of customs.) (D 9; GS 117) Under this pressure, the feeling of obligation to obey customs approaches a categorical nature.

This is a significant step in the development of morality and the feeling of guilt. Note first that the perceived threat of punishment, and anxiety over the same, is responsible for making religious obligations into a matter of ethical concern. Failures to meet one's religious duties, understood as a potential cause of harm to the community, are now understood as wrongs to the community, and so invite punishment. As religious practices and beliefs become more sophisticated, they become increasingly a matter of living up to some ideal and embodying the virtues that are favored by the gods, or God. But the felt connection between the ethical concerns of the community and the personal concerns of living up to an ideal remains (however disconnected these concerns are in fact from each other), and continues to influence the development of morality. Eventually, there is little to no distinction made between an ethical failure and wrong, and a failure in personal virtue, which is felt as wrong and as inviting punishment.

This conflation between the ethical and the personal is a central feature of the underlying conditions that make the interpretation of guilt possible. Through this conflation, because personal failures are so often private matters, the individual will feel a sense of personal unworthiness (as opposed to the feeling of corporate unworthiness that a community might feel with regard to pleasing the gods). And since this is a private feeling, once the interpretation of guilt is offered and accepted, the individual will have sole responsibility for that guilt and a pretext to engage in self-punishment.

In the relationship between buyer and seller, debt is material and finite and can be definitively paid off. In the relationship to the community, following the rules is sufficient to repay one's debt. Failures in either of these relationships have their equivalences in punishment, which serves to repay the debt. But the gods do not appear to be satisfied by mere obedience to customs. Following the rules of the community might be a necessary condition for pleasing the gods, but it is not a sufficient condition. One must go further, one must offer the gods sacrifices.

But anxiety, and suspicion that repayment has not been adequate, produces increasing emphasis on the need for greater sacrifices. A sacrifice is not genuine unless it involves loss and suffering. Thus, the relationship to the gods makes suffering not just one means of providing repayment, but the preferred, and perhaps only, means.¹⁴

¹⁴ This idea will be important for understanding the Greeks and the Jewish community prior to the Exile. An emphasis on sacrifice serves a covert purpose. Because it involves a voluntary act that produces suffering it effectively internalizes aggression and prevents the development of the bad conscience.

Thus, the relationship to the gods produces a feeling that obeying customs is both imperative and inadequate to please the gods. As we will see in the next chapter, under certain conditions this fosters a sense of futility and impotence that produces resentment and leads to the bad conscience.

It should also be noted that the relationship to the gods is in tension with the developments we saw in the relationship between the individual and the community. In that relationship, the community's increased sense of its own strength decreased the severity of its response to crime. The community was able to understand some crimes as less serious than others, and punishment, became milder. Criminals were protected against general anger, and punishment was limited to equivalences established by law.

This development appears to be part of Nietzsche's understanding of the logic of the creditor-debtor relationship. He claims that, "The 'creditor' has always become more humane to the degree that he has become richer; finally the amount of injury he can bear without suffering from it even becomes the *measure* of his wealth." (GM II:10) Nietzsche goes on to claim that the justice that begins with "everything can be paid off, everything must be paid off" (i.e. the equivalence principle) ends with "letting the one unable to pay go free." He calls this the "self-cancellation of justice," also known as "mercy," which is the "privilege of the most powerful." Moreover, recall that Nietzsche claims that the pleasure of being allowed to punish is inversely correlated with the creditor's social standing, the lower and baser the creditor's standing, the greater his pleasure in making suffer. (GM II:5)

Put simply, creditors with a sense of power (whether an individual or the community) feel less harmed and are less angered by injury, and can afford to be more magnanimous and merciful towards the agent of that injury. By contrast, those who lack power feel more harm and more anger, feel greater need to make suffer, and take greater pleasure in doing so.

Paradoxically, while each success of the community provides evidence of the gods' power, they are not seen as more magnanimous and merciful creditors, who could dismiss certain losses, but rather, as demanding greater payments and threatening greater punishment for failure to do so. Thus, the gods appear to be a reflection of those who lack power, feel more harm over unpaid debts, more anger, and the desire to make suffer.

So while the community's sense of its own power results in more leniency towards the debtor, the community's sense of the gods' power does not produce the same result. As the gods' power increases the community's debt increases as well. Thus, the community's relationship to the gods has ceased to follow the original logic of the creditor-debtor relationship.

My argument in the next chapter will be that it is primarily the sense of overwhelming debt to the gods, that cannot be repaid, that is the source of the feeling of guilt. Once the hunter-gatherer tribes that we have examined in this chapter are conquered and subjugated into the first politically organized community (the state), and undergo internalization of their aggressive instincts, their perceived debt to the gods will become a source of systematic frustration, and sense of worthlessness, that produces

resentment and gives birth to the bad conscience and guilt. Moralized guilt, once it develops, will share the same features that we found here in debt consciousness: anxiety over punishment, doubt over one's ability to fulfill one's obligations, the tension between ethical duties as both imperative and inadequate to please God, and a consuming sense of worthlessness.

Chapter Two: The Bad Conscience and Guilt

“Oh, this insane sad beast man! What ideas occur to it, what anti-nature, what paroxysms of nonsense; what *bestiality of idea* immediately breaks forth when it is hindered only a little from being a *beast of deed!* ...”¹⁵

Introduction

As we know, Nietzsche claims that the feeling of moral guilt has its origins in the material concept of debt. We have seen that the creditor-debtor relationship between buyers and sellers involved contractual relationships and employed punishment to enforce those contracts and “promises.” It is in this context that the psychological connection between debt and suffering and the idea that an equivalent amount of suffering can be determined for every debt is established. And this stage, however, to owe a debt is only to be “legally liable” to proportionate suffering, for purely economic reasons.

Over time, the psychological connection between “promising” and “punishment” develops into a primitive feeling of “obligation” to treat promises as duties that must be fulfilled. When the creditor-debtor relationship is interpreted onto the individual’s relationship to the community and the ancestors, these relationships are understood as contractual, and the individual is understood to owe obedience to the community and sacrifice to the gods. Breaches of custom (either ethical or ritual) are understood as violations of the contract that must be repaid through punishment. Thus, the individual develops a feeling of debt consciousness to both the community and the gods.

¹⁵ GM II:22

These developments occurred long before the first genuinely political communities. The founding of the first political “state” affects these already existing features of human psychology, and creates the conditions that lead to the development of the feeling of guilt.

The issue now is to trace out the connections between the effects produced by the conditions of the state on human psychology that lead to the feeling of guilt. Nietzsche mentions a number of factors that must be taken into account: loss of freedom and the internalization of the instincts; the conscience; the animal bad conscience; the active bad conscience; resentment; animal cruelty and the will to power; the pressure of unpaid debt to the gods; and the moralization of guilt.

Providing an account of guilt, however, is further complicated by the fact that it cannot be completed without taking into account factors that Nietzsche does not explicitly describe or explain: the influence of noble morality and religion; sovereign individuality; the role of the primitive instinct to obedience and “formal conscience”; and the role of the priests and the ascetic ideal in providing the interpretation that produces guilt.

The purpose of this chapter is to present and explain Nietzsche’s genealogical account of the development of guilt. My interpretation involves a number of distinctions. Some of these are uncontroversial, others will be defended along the way.

When Nietzsche first mentions the “bad conscience” it is juxtaposed with another “gloomy” thing: the conscience. According to my account, these are not separate

faculties, but rather, the bad conscience is an interpretation of a conscience that has been distorted in several significant ways, and is in conflict with itself.

The bad conscience is clearly synonymous with “consciousness of guilt.” To keep matters clear, further distinctions will need to be made. The bad conscience which is the consciousness of guilt, is distinct from the “animal bad conscience,” which does not involve guilt, but is descriptive of an animal whose aggressive instincts have been repressed, internalized, and are vented against the animal itself. The consciousness of guilt that is the bad conscience is a pervasive feeling of guilt that is produced by an interpretation of the animal bad conscience, and is what Nietzsche calls the “moralization of guilt.” According to my interpretation, this pervasive guilt, which is the product of Judaism and Christianity, is the target of Nietzsche’s objections.¹⁶ I am concerned to defend Nietzsche from the charge that his opposition to guilt implies that he would eliminate from human psychology any painful feeling over having done wrong.

Accordingly, I distinguish between the pervasive guilt that Nietzsche opposes and what we might call “episodic guilt,” or “locally reactive guilt” as some interpreters prefer. I

¹⁶ Aaron Ridley claims that the internalization of the instincts is solely responsible for the bad conscience, and involved no transcendental presuppositions or the concept of God. On his account the slaves invent the bad conscience to provide an outlet for their repressed cruelty. (Ridley, 1998, p. 37)

Bernard Reginster claims that Christianity distorts the ordinary feeling of guilt. On his account the bad conscience develops when human beings feel an obligation to live up to ethical standards and engage in self-reproach over failure to do so. Debt consciousness is transformed into guilt when these failures are felt to diminish the person’s self-worth. Christianity distorts this ordinary feeling by turning it into a “rational passion” for self-condemnation. (Reginster, 2011, p. 57)

Matthias Risse claims that the bad conscience as a feeling of guilt arises from an earlier form of the bad conscience that involves a sense of indebtedness to the gods. There are terminological differences that make it difficult to be certain, but it seems that my account is closest to Risse’s and might be seen as providing a further historical development of his view. (Risse, 2001)

will argue that “episodic guilt” is a feature of a healthy conscience, and that Nietzsche has no objection to episodic guilt, to the extent that it is not connected to, or influenced by, pervasive guilt.

My interpretation differs from other interpretations. I claim that the feeling of pervasive guilt is ultimately the product of religious developments (influenced by economics), that occurs when the slave population of the state inherit noble value terms and adopt the gods of the nobles. This implies that these slaves were (at least proto-) sovereign individuals, whose sovereign individuality was corrupted by the interpretation of their suffering as guilt.

Subjugation into the conditions of the state forced the “internalization” of the slaves’ aggressive instincts through denying them external expression. The pressure to adapt to new social conditions, political structure, and new customs and rules forced the slaves to redirect their aggressive instincts towards denying the expression of any instincts that would lead to disobedience to the rules and customs of the state. Internalized aggression became a will to power over the other instincts in connection with the instinct to obedience. This allowed the slaves to overcome their instinctive impulses and conform to the customs of the state. The hierarchy between the will to power and the instinct to obedience over the the other instincts provided satisfaction to both instincts. The instinct to obedience kept the slaves from debt and punishment and the will to power was satisfied through aggressively dominating the other instincts. The relationship between

the instinct to obedience and will to power, made instinctive, constitutes the conscience of the slaves, which implies that they were (again, in some sense) sovereign individuals.

As sovereign individuals the slaves were capable of self-regulating behavior in light of existing norms and ideals, without requiring the threat of external punishment to enforce obedience. The only ideals available to the slaves, however, were provided by noble culture. The slaves inherit the concepts of “good and bad” from the nobles and adopt the nobles’s gods. In doing so they inherit the unpaid debt to the nobles’ gods, the idea that the debt is repaid through living up to noble ideals, and the judgment that they themselves are “bad” and are thus incapable of living up to those ideals and repaying their debt. The result is an inescapable consciousness of debt, unfulfilled obligation, and fear of punishment. This produces suffering, a frustration of the will to power (further suffering), and resentment.

Ressentiment inspires the invention of the bad conscience as an interpretation of the slaves’ predicament. They are impotent to keep their obligation to repay debt, and feel unable to avoid punishment and suffering. The ascetic priest interprets this condition as “guilt before God”, and “punishment for sin,” and attaches this guilt to the slaves’ animal instincts, as what lead them to sin against a holy, and transcendent God. The feeling of debt consciousness becomes the feeling of guilt over sin, and the suffering produced by the animal bad conscience and resentment becomes God’s punishment for sin. This is the “moralization” of guilt, which identifies animal nature as the source of moral failure. Moralization is significant because the identification of animal nature as the source of

evil, or wrongdoing, can be made increasingly more abstract to become “selfishness,” or “egoism,” and can continue to influence human psychology absent any belief in God.

Thus, my thesis is that moral guilt is the product of two main influences: economics and religion. Economic practices established a connection between debt and punishment. The threat of punishment causes discharging debts (in whatever form) to be felt as an obligation. When this feeling of obligation becomes instinctive, the connection between debt and punishment is made instinctive with it, and the person’s conscience produces painful feelings over the thought of moral failure, whether this is to an individual, the community, or the gods. In the case of other individuals and the community, failure to meet one’s obligations is typically met with punishment, and the debt is repaid. But the ascetic priest presents the slaves with an unattainable ideal. Since there is no one, other than the gods, to punish failure to attain the ascetic ideal, the slaves must either punish themselves or live under the pressure of ever-increasing debt consciousness. Thus, it is the relationship to the gods (or God) that produces the overwhelming sense of indebtedness, impotence, and suffering that is interpreted by the priest into guilt. The ascetic ideal produces the idea of a holy, transcendent God and identifies the animal instincts as the source of sin and moral failure. Human beings fail to live up to this ascetic ideal by willfully acting on their animal instincts. The connection between animal nature and moral failure survives the death of God, when specifically religious and ascetic ideals are replaced by secular versions, such as: selflessness, self-sacrifice, and altruism, which

allows the feeling of pervasive guilt – personal failure to live up to an ideal standard – to survive without God.

2.1 The Influence of Debt Consciousness

Nietzsche claims that the feeling of guilt has its origins in the creditor-debtor relationship and debt consciousness. But the creditor-debtor relationship was established in the earliest human communities, and guilt does not develop until after the end of those communities and the beginning of the state. Thus, guilt is the product of the way that debt consciousness is affected and transformed (one might say corrupted), by factors that are created through the establishment of the state.

We should recall a few facts about the the earliest human communities before we proceed with an examination of the bad conscience and guilt. These communities were responsible for producing an “instinct for obedience,” or “formal conscience,” in the members of the community. The morality of custom and the practice of banishment for breaking customs preceded the development of the creditor-debtor relationship, and were responsible for conditioning human beings to remember and follow the basic rules of the community. Thus, it created a primitive sense of obligation (however much this might be motivated by self-interest).

The creditor-debtor relationship evolves to meet the demands of growing communities that began to engage in more complex transactions of goods or services, that could no longer be conducted through immediate barter and exchange. The creditor-debtor relationship is based on making agreements and so results in a situation where

promises must be taken as sources of obligation. It accomplishes this by formalizing promises through a contract, and employing punishment as both an incentive for the debtor to take the promise seriously, and a means of providing the creditor with repayment through the satisfaction of causing the debtor to suffer.

The creditor-debtor relationship introduces several new developments to the already existing formal conscience. First, it serves to impress promising, as a source of obligation, on the formal conscience: keeping promises is now one of the rules of the community and to be remembered. Second, it introduces the idea of voluntary obligations – obligations that the individual incurs through an agreement. Lastly, it establishes both a psychological connection between indebtedness and suffering, by making punishment a means to repay a debt, and introduces the idea that there is always an equivalence to be found between a debt owed and a commensurate amount of suffering.

The economic thinking of the creditor-debtor relationship is used to interpret the relationship of the individual to the community. The individual is understood as having made an agreement with the community in order to live within its protection. The community has advanced goods to the individual (such as protection), so the individual stands as a debtor to the community as creditor. Repayment of this debt comes in the form of obedience to the rules and customs of the community. Rule-breaking is interpreted as a violation of the terms of the contract, which requires punishment as the appropriate means of repayment. Thus, the individual develops a sense of debt

consciousness towards the community in connection to the primitive feeling of obligation (in a purely prudential sense, at this stage) to follow its rules.

The creditor-debtor relationship is also applied to the community's relationship to the ancestors, or (over time) gods of the community. The ancestor-gods are understood to be responsible for the founding of the community, and for continuing to give it aid through their continued presence as powerful spirits. Thus, the ancestors or gods have provided the community with goods in advance and stand as creditors to the community as debtor. Repayment of the debt comes in the form of following the customs established by the ancestors and providing them with offerings and sacrifices. Thus, the individual, as part of the community, participates in a collective sense of debt to the ancestors.

While the idea that suffering can repay a debt is already questionable (whether to another individual or the community) it is the relationship to the ancestors and gods that turns debt consciousness into a feeling that is susceptible to being interpreted as guilt. As noted in the previous chapter, the relationship to the ancestors differs from the other forms of the creditor-debtor relationship in several notable ways.

The ancestors are not present to the community such that they could make their demands or desires known, so cannot be seen as satisfied with what they were offered. This leaves the form and amount of repayment (sacrifices) to the fearful imagination of the community. Fear of the gods introduces the suspicion that they have not been properly repaid. Every success of the community increases the feeling of debt and the growing suspicion that it has not been properly repaid, such that it periodically requires

an “enormous counter-payment” through blood and sacrifice. Thus, the feeling of debt to the gods involves anxiety and doubt concerning the means to repay them.

Since the gods are understood as responsible for the community’s continued success, collective anxiety over providing repayment places more pressure on the individual to follow the customs established by the gods. The individual’s failure to follow customs becomes a serious matter, since it invites the displeasure of the gods and threatens the entire community. This leads to an increasing emphasis on following customs that is felt by the entire community. Nietzsche goes so far as to claim that the community feels a collective responsibility for the individual’s violation of customs. Under this pressure, the feeling of obligation to obey customs approaches a categorical nature.

Anxiety and suspicion that repayment has not been adequate also shifts the means of repayment from following customs – in the sense of the rules of the community established by the gods – to sacrifice and suffering. In the relationship between buyer and seller, debt is material and finite and can be definitively paid off. In the relationship to the community, following the rules is sufficient to repay one’s debt. Failures in either of these relationships have their equivalences in punishment, which serves to repay the debt. But the gods do not appear to be satisfied by mere obedience to customs. It is not sufficient to be “good” – in the sense of being a good member of the community – one must go further and offer the gods sacrifices. A sacrifice is not genuine unless it involves loss and suffering. Thus, the relationship to the gods makes suffering not just one means of providing repayment, but the preferred and perhaps only means.

All of these debt relationships helped to establish a deep psychological connection between debt and suffering in the members of these communities. This connection between debt and suffering is the source of the primitive feeling of obligation. The creditor-debtor relationship is much earlier than the founding of the state, so the earliest subjects of the state brought with them already existing feelings of indebtedness to the community and the gods. The effects of being socialized into the political structure of the state alters both of these relationships and places them in conflict with each other, ultimately leading to the feeling of moral guilt.

2.2 The State: Conquest and Subjugation

Nietzsche claims that the bad conscience is a result of a dramatic development in human history: the foundation of the oldest political “state.” He describes the conquest, subjugation, and enslavement of nomadic communities organized according to blood relationships (GM II:20), by a smaller, but aggressive, and more militarily-organized clan of warriors. This clan will eventually be known as the nobles, so I will refer to them as such, though it should be kept in mind that it takes some time for the concept of nobility to develop. Accordingly, I will refer to the conquered nomads as the slaves.

The state, Nietzsche claims, is founded by: Some pack of blond beasts of prey, a race of conquerors and lords, which, organized in a warlike manner and with the power to organize, unhesitatingly lays its terrible paws on a population enormously superior in numbers perhaps, but still formless, still roaming about. (GM II:17)

Since the state began through an act of force, the subsequent structuring of the population had to continue as a matter of force. Thus, the political structure that emerged was not a

matter of agreement or social contract. “Whoever can give orders, whoever is ‘lord’ by nature, whoever steps forth violently, in deed and gesture — what does he have to do with contracts!” (GM II:17) The oldest state:

Accordingly made its appearance as a terrible tyranny, as a crushing and ruthless machinery, and continued to work until finally such a raw material of people and half-animals was not only thoroughly kneaded and pliable, but also formed. (GM II:17)

Here we see the nobles’ “power to organize” in action. Their work is “an instinctive creating of forms; impressing of forms,” which they apply to the conquered slaves. The end result of impressing form on this originally formless population is:

A ruling structure that lives, in which parts and functions are delimited and related to one another, in which nothing at all finds a place that has not first had placed into it a “meaning” with respect to the whole. (GM II:17)

The slaves are incorporated into the nobles’ project of building and maintaining the state. They are forced into new, unfamiliar roles and functions, whose purposes are dictated by the nobles. New roles and activities introduces new customs, rules, and expectations on the slaves. This, Nietzsche claims, is the “most fundamental of all changes” that human beings have experienced. The pressure from this change causes them to fall into the “deep sickness” that is the bad conscience.

The lives of the slaves in their previous existence as nomadic peoples had them “happily adapted to wilderness, war, roaming about, adventure.” But with their subjugation they find themselves “enclosed once and for all within the sway of society and peace.” (GM II:16) The slaves are subjected to an “oppressive narrowness and

regularity of custom,” which prohibits them from engaging in the activities that were part of their previous life.

Nietzsche claims that the transition to political society was “not gradual, not voluntary,” and not “an organic growing into new conditions,” but rather “a break, a leap, a compulsion, and inescapable doom.” This feature of Nietzsche’s account is as important for what Nietzsche does *not* say (but leaves for us to uncover), as for what he explicitly states. What he does *not* discuss explicitly are cases where the transition into political society is voluntary, gradual, and an organic growth, and the effect that this might have on the development of the conscience. We know that the nobles do not develop the bad conscience. Describing the nobles’ unconscious artistry of imposing form on the slaves, he says, “They are not the ones among whom ‘bad conscience’ grew, that is clear from the outset...” (GM II:17) But, since Nietzsche describes the nobles as “inventive in consideration, self-control, tact, loyalty, pride, and friendship,” (GM I:11), who describe themselves as “the truthful,” and even have reverence for their enemies (GM I:10), they are clearly capable of a robust ethical life (at least amongst equals), which requires the possession of a conscience. If this is the case, it implies that having a genuine ethical conscience does not require the features that are found in the bad, or guilty, conscience.

2.2.1 Internalization of the Instincts

Nietzsche does say that the forced, sudden, and involuntary transition into political society alienated the slaves from their own dominant instincts. These instincts were “their old leaders, the regulating drives that unconsciously guided them safely” in an

environment involving hunting, war, wandering, and adventure. But, through sudden, forced enclosure in the conditions and requirements of the state “all at once all of their instincts were devalued and ‘disconnected’.” Nietzsche describes the slaves as feeling much the way that former water animals must have felt when forced onto the land to survive. Their dominant instincts cannot be used to navigate their new environment, and are liabilities, since the nobles used the “terrible bulwarks of the state,” which include punishment especially, to prevent the expression of the aggressive instincts. (The nobles would certainly punish any aggression directed towards them, but it seems that their interest in building a state would also require them to punish acts of aggression between the slaves.) This leaves the slaves to rely on their “poorest and most erring organ”: conscious “thinking, inferring, calculating, connecting cause and effect” to navigate their social environment. (GM II:16)

The problem, however, is not simply a matter of learning new skills, or becoming proficient with underdeveloped ones (like thinking). The problem is that the aggressive instincts still exist and continue to make demands, even while they are denied external expression. According to Nietzsche, “all instincts that do not discharge themselves outwardly turn themselves inwards.” (GM II:16) This is the internalization of the instincts. Internalized, the aggressive instincts – the “instincts of the wild free roaming human...hostility, cruelty, pleasure in persecution, in assault, in change, in destruction” – are turned against the self and vented on the self. Nietzsche likens the effects of internalization on human beings to the behavior of an animal that “beats itself raw on the

bars of its cage.” Taken out of the wild, the human animal tears into itself, persecutes itself, and mistreats itself, in an attempt to create a “wilderness within the self”: a place of adventure, torture, and danger, where the aggressive instincts can find expression and satisfaction.

2.2.2 The Will-To-Power

The instincts of the slaves just mentioned (hostility, cruelty, etc.) are what Nietzsche calls the “instinct for freedom,” which he also claims is synonymous with the will to power. The will to power is central to the development of the bad conscience, and has been defined by scholars in a variety of ways, so it must be addressed before we proceed.

Reginster claims that the will to power is “the self-standing desire to be an effective agent, that is to say, an agent who effectively bends the world to his will (his values). (Reginster, 2018, p. 13) Reginster rejects the idea that the will to power is essentially the desire to inflict cruelty or make suffer, rather, the will to power seeks satisfaction through overcoming resistances. (Reginster, 2011, p. 60) I find this implausible for several reasons.

In the first place, as Avery Snelson pointed out,¹⁷ once punishment becomes a practice in the creditor-debtor relationship, and continues to be used as one of the “terrible bulwarks” of the state, an important development takes place. The actual execution of punishment is taken out of the hands of the injured party and carried out by the “authorities.” It is hard to imagine how the injured party, who demands satisfaction

¹⁷ Conversation with Avery Snelson, Winter 2018.

through punishment, experiences the feeling of “overcoming resistance” when the authorities carry out punishment on a captive prisoner. But it is easy to imagine how the injured party might enjoy the spectacle of cruelty and seeing-suffer.

Moreover, even if the will to power can be satisfied through overcoming resistance, the lives of early human beings, living in a world of hunting, war, and adventure, would have primarily satisfied the will to power through aggressive, violent actions. Thus, there would be a strong psychological connection between making suffer and satisfaction of the will to power, which can be aptly described as cruelty.

Finally, Reginster appears to overlook Nietzsche’s numerous claims appeal to the pleasures of cruelty, however sublimated through art, religion, or culture. Even the most scholarly search for knowledge, Nietzsche claims, always contains a drop of cruelty, in that it is committed to finding the truth at the expense of one’s own cherished beliefs or illusions.

Maude Clark argues that will to power is a second-order drive of the drives. The drive itself (such as hunger) has a target (food) and the aim to acquire food. But the hunger-drive also has a second-order aim to be the drive that dominates the organism and takes precedence over the demands of other drives. In the case of the instinct for freedom, Clark argues that it is an instinct to resist being restricted. But when conditions demand that a particular instinct be restricted in expression (such as aggression, in the context of the state), that instinct is directed at having power over other instincts. In this case, the will to power of the drive turned within seeks satisfaction by denying satisfaction to other

drives, and becomes a will to control and command. In the case of the slaves, their strongest, dominant instinct is aggression, so aggression seeks power by denying the expression of the other drives, and will be a central influence on the development of the bad conscience. (Clark & Dudrick, 2012, pp. 205-207)

When we turn to consider the sovereign individual, Reginster's notion of a "self-standing desire to be an effective agent" may be helpful, since I will argue that the bad conscience is a distortion of sovereign individuality, such that the slaves of bad conscience have a standing desire to effectively engage in self-punishment. But I will maintain that this project is rooted in the pleasure of cruelty. In the main I largely follow Clark's interpretation of the will to power because it most accurately illuminates the developments that take place through the internalization of the instincts that lead to sovereign individuality and leave the human being open to the bad conscience.

2.2.3 The Animal Bad Conscience

My account of the bad conscience distinguishes between the bad conscience in its beginnings, which I call the "animal bad conscience," and the bad conscience in its final form, which Nietzsche at one point calls the "active bad conscience," but I will typically refer to as simply the "bad conscience." I will present the animal bad conscience and explain why I make this distinction, before preceding to the more difficult project of providing an explanation for the fully developed bad conscience.

Nietzsche claims that when the state banishes the expression of an enormous amount of freedom, the instinct for freedom turned within and "forcibly made latent..driven back,

suppressed, imprisoned within...discharging itself only on itself...is bad conscience in its beginnings.” (GM II:17) This reemphasizes the point he made earlier when he claimed to be presenting the *origin* of the bad conscience (GM II:16), and is echoed again in his description of the person attempting to create a “wilderness within,” where he says “this fool, this longing and desperate prisoner *became* the inventor of ‘bad conscience’.” (GM II:17) [my emphasis] The use of the word “became” further suggests that the invention of the bad conscience is yet to come and is not to be equated with the mere internalization of the instincts. Nietzsche describes the “active bad conscience,” as a:

Secret self-violation, this artists’ cruelty, this pleasure in giving oneself a will, a critique, a contradiction, a contempt, a ‘no’; this uncanny and horrifying-pleasurable work of a soul-compliant-conflicted with itself, that makes itself suffer out of pleasure in making-suffer, this entire active bad conscience... (GM II:18)

In this passage we see that the “active bad conscience” involves a focussed attempt to impose form on oneself that includes negative self-evaluation. Thus, the mere internalization of the instincts against the self is not sufficient for the bad conscience in its final form. The internalization of the instincts leads them to seek new ways to find gratification and causes them to grow stronger in their struggle to do so. Thus, the instincts that are most denied external expression, eventually become the strongest.

This is crucial to the development of the bad conscience, but it is only the beginning of that development. Before the instincts can learn new ways to find gratification, they are forced to vent themselves on each other. This is the “animal bad conscience,” – an involuntary reaction to new, disorienting, and unfavorable conditions for the expression

of the dominant instincts. But, the “animal bad conscience” only indicates an animal that suffers because it can no longer act on its own dominant instincts, despite their continued demands for satisfaction.

The slaves suffer from internalized aggression vented within, but they also suffer anxiety over punishment. Their dominant instincts lead them to violate the new rules and customs of the state, which leads to punishment, and produces mistrust of the dominant instincts. Since the slaves are brought into the state with a sense of debt consciousness towards the community, the continued demands of their dominant instincts become associated with violation of the rules and the prospect of punishment. Thus, the slaves live in anxiety over punishment for inadvertently breaking rules because of their instinctive impulses. The result is an animal that is at odds with itself and lashes out against itself with unfocused hostility. Again, this is only the animal bad conscience. The development of the “active bad conscience” will require further changes to the human animal.

2.3 The Instinct to Obedience and the Instinct for Freedom

The bad conscience is initiated through the internalization of the instincts, which turns them against each other and produces the animal bad conscience. As I have already stated, internalization alone is not sufficient to produce the bad conscience – that requires other important developments. But it is important to have a clear picture of the struggle between the instincts that incorporates these other developments to produce the bad conscience. In this section I present the underlying internal conflict that occurs within the

slaves and the tension produced through the formation of the state between two primitive instincts: the instinct to obedience and the instinct for freedom. I claim that it is the relationship between these two primitive instincts that ultimately produces a hierarchical relationship between the slaves' drives and provides the basis for the bad conscience.

The creation of the state thrusts former nomadic people into slavery and forces them to follow new rules and customs. The slave must adapt quickly to new social conditions, in order to survive. The instincts that provided them with an evolutionary advantage in the wild are now a source of liability in the context of society and peace. They lead the slaves to violate rules and customs, which incurs debt, and leads to punishment. In order to avoid punishment, the slaves need to gain mastery over their aggressive instincts.

Nietzsche claims that mastery over some instinct is only a sign that another instinct has become dominant (BGE 117; D 109). But Nietzsche also claims that the dominant instincts of the slaves were devalued and disconnected. And so they were.

Except one. As we saw in the discussion of the creditor-debtor relationship, life in the nomadic communities prior to the state, involved obedience to rules and customs. These were enforced through harsh sanctions that created a memory of the basic rules that are the "requirements of social co-existence." This selected for, and reinforced, the herd instinct of the members of the community, producing an instinct to obedience, or "formal conscience." Nietzsche claims that the formal conscience is a need to obey that is made innate, and which "seeks to be satisfied and to fill out its form with a content." (BGE

199) Thus, we can say that the formal conscience provides a primitive sense of obligation or duty, which is to remember and follow the rules and customs of the community.

In the context of the state, the instinct to obedience would drive the slaves to find a way to conform to the rules of their new community. They would be driven to satisfy their innate need to obey by filling out the “formal conscience” with this new content. To do this, however, they must find a way to remember the new rules, and prevent their dominant instincts from interfering with this.

Nietzsche claims that human instinct “intuited in pain the most powerful aid of mnemonics.” “One burns something in so that it remains in one’s memory: only that which does not cease to give pain remains in one’s memory.” (GM II:3) The mnemonics of pain were employed to produce the formal conscience, and the memory of promises as a source of obligation, in the creditor-debtor relationship. There is no reason to think that the slaves have lost this intuition connecting pain and memory. Thus, the slaves take sides against themselves and redirect their aggressive instincts against the external expression of the instincts, including, especially aggression. They declare war on all of the old instincts that have been devalued, disconnected, and denied external expression, by using aggression and cruelty against themselves to conform to their new situation.

According to Clark’s theory of the drives, the restriction placed on the aggressive drives (the instinct to freedom) leads them to seek their own expression by denying the expression of other drives. They seek to control and command the other drives and instincts, and when this occurs the instinct for freedom is a will to power. (Clark, 2012)

What we see in the slaves initially, then, is a conflict between the instinct for freedom and the instinct to obedience. In the context of their previous nomadic lives, there was no significant conflict between these instincts. As I have argued, natural selection favored communities whose members had a sense of reciprocity towards each other, as well as aggressive instincts that are conducive to survival. The practice of banishing non-obedient, non-reciprocating members of the community accomplishes this. Further, the aggressive instincts were allowed expression in the wild: in the hunt and in fighting enemies. Any tensions that arose from the demands of obedience within the community, would be discharged outside the community, in much the same way that Nietzsche describes the nobles in GM I:11. Thus, we can add to Nietzsche's claim that the nomads were happily adapted to life in the wilderness, the claim that they were happily adapted to life in their communities.

In the context of the state, however, this is not the case. The instinct to obedience is now directly opposed to the instinct for freedom – at least with regard to actions within the community and obedience to customs. Restricted and turned within, the instinct for freedom becomes a will to control the other drives, and finds satisfaction through denying any expression of those drives that would lead to breaking customs. And in doing so, the will to power satisfies the instinct to obedience by enforcing obedience to customs.

So, what we find is an “alliance” between drives that results in a hierarchical structure of the drives. The instinct to obedience is given authority to rule over the other drives, because it provides the best (and in this context, only) satisfaction of the dominant

aggressive instinct's will to power. The expression of all other drives is subordinated to the aggressively enforced demand to conform to customs. Here we should note that aggression is employed not only to "master" and "overpower" the expression of other instincts, and obedience to customs, but also, ironically, its own external expression, which in turn satisfies aggression and promotes it to a position of authority. This hierarchical arrangement that connects obedience and aggression provides the basis for the development of the conscience.

2.4 Necessary Conditions for the Bad Conscience

Nietzsche does not explicitly spell out how the animal bad conscience develops into the bad conscience, but he provides us with clues that will allow us to pull together an explanation. What we will find is that the bad conscience is a product of resentment, the activity of giving form to oneself through the will to power, negative self-assessment, and the inheritance of debt from the gods. In this section I will piece together these clues to provide an explanation of the development of the bad conscience.

The first clue is Nietzsche's claim that the "human being of resentment" has the "invention of the bad conscience on his conscience." (GM II:11) This clearly indicates that the bad conscience is a product of resentment. But we should pause to note an important distinction. Resentment can be episodic or it can be a stable feature of a person's psychological make-up. In episodic form, resentment is a response to a situation that temporarily frustrates one's will to power. In its episodic form, even the

nobles are capable of feeling resentment, though with them, it typically “runs its course and exhausts itself in an immediate reaction” such that it does not endure.

Nietzsche claims that the “human being of resentment” is responsible for the bad conscience. This indicates that resentment is a stable feature of that person’s psychology. Resentment, in this sense, is produced when a person’s will to power has been systematically and prolongedly frustrated, leading to a pervasive sense of impotence, anger over that impotence, and the desire to express one’s will through retaliation. Nietzsche describes prolonged resentment as festering and poisoning the person, producing a desire for revenge against whatever is taken to be the source of that frustration. (Since episodic resentment is not central to this account, I will drop the locution “human being of resentment” and simply use “resentment,” unless otherwise indicated.)

If the bad conscience is a product of resentment, then it is subsequent to the development of resentment as a widespread feature of at least some instances of human psychology. Thus, any account of the bad conscience must explain the development and source of resentment.

Another clue is found in Nietzsche’s claim that the instinct for freedom (the will to power) motivates the nobles to impose form on the state and the slaves, and the slaves to impose form on themselves through internalized aggression. Both are described in artists’ terms. The nobles are the “most involuntary instinctive artists there are,” whose work in creating the state is ruled by a “terrible artists’ egoism without guilt,

responsibility, or consideration, that knows itself to be justified in its work.” (GM II:17)

The slaves are described as engaged in the same work, but within themselves, on a smaller and pettier scale. Nietzsche describes the slaves as enjoying the “artists’ cruelty” of giving themselves form, of burning into themselves “a will, a critique, a contradiction, a contempt, a ‘no’.” (GM II:18) This “horrifying-pleasurable work of a soul compliant-conflicted with itself, that makes itself suffer out of pleasure in making-suffer” is the entire “active bad conscience.” Thus, it is the will to power turned within that gives form to the self, and “creates for itself the bad conscience and builds negative ideals.”

This passage implies that two further developments must take place in order to produce the bad conscience: the slaves must take up an ideal against which they judge themselves negatively, and they must begin to engage in the same form giving activity as the nobles.

Nietzsche tells us that the slaves inherit a “propensity for establishing forms of rank” from the nobles. (GM II:20) The slaves may have employed this propensity within their own communities, but of greater importance is the way that it is applied to their own instincts through their own negative self-assessment, that explains the development of the bad conscience.

Establishing orders of rank implies that there is some standard against which things are ranked. Nietzsche tells us that the slaves inherit this standard from the nobles. But, what he tells us is rather surprising after having read GM I. Nietzsche claims that the slaves inherit the concepts “good and bad” from the nobles. Reading Nietzsche’s

description of noble and slave morality in GM I can give the impression that the slaves always felt resentment towards the nobles and only ever thought in terms of the “good and evil” of slave morality. But Nietzsche is clearly suggesting that before that occurred, the slaves became acclimated to noble culture and thought in terms of “good and bad.” This is an under-appreciated feature of Nietzsche’s account, and I take it be central to the development of bad conscience, because it provides the best explanation for the development of resentment.

An important, and frequently overlooked, clue is found in the same passage that mentions the slaves’ inheritance of “good and bad,” and the psychological propensity for establishing orders of rank:

...in the same way that it inherited the concepts “good and bad” from the clan nobility (together with its propensity for establishing orders of rank), humanity also inherited, along with the deities of the clan and tribe, the pressure of the still unpaid debts and the longing for the redemption from the same. (GM II: 20)

The mention of “clan and tribe” could suggest the gods that the slaves worshipped prior to their subjugation by the nobles. But it is important to remember that the nobles were originally nomadic hunter-gatherer groups that can be described as clans or tribes. If there is any doubt that debt mentioned here is the debt owed to the new gods of the state – the gods of the nobles – Nietzsche goes on to say, in a parenthetical comment:

The transition is made by those broad slave and serf populations who adapted themselves to the cult of the gods practiced by their lords, whether through force or through submissiveness and mimicry: starting from them, this inheritance then overflows in all directions. (GM II:20)

There are several features of this passage that have not been given due treatment in the past. The first is that the slaves incorporate the gods of the nobles into their own practices. (Whether this is in addition to their own tribal gods, or at the exclusion of those gods, or some form of syncretization with their old gods is immaterial.) This further indicates that the slaves are being increasingly acclimated to noble culture.

The claim that the slaves inherit the debt to the nobles' gods, and that this debt "overflows in all directions, is even more significant. It implies that the slaves continue to think of their relation to the gods in terms of the creditor-debtor relationship, and that their sense of debt consciousness is overwhelmed. Moreover, the passage suggests that the concept of debt will now be applied to new things – perhaps everything.

Nietzsche's claims about debt alone should indicate that an important development has taken place. But what Nietzsche says next reveals that the slaves' inheritance of debt is more than just an important development, it is a transformative one:

For several millennia the feeling of guilt toward the deity did not stop growing and indeed grew ever onward in the same proportion as the concept of god and the feeling for god grew on earth and was borne up on high. (GM II:20)

This is the first mention of "guilt" in the aphorisms that deal with the creditor-debtor relationship to the gods. Until this point, Nietzsche has only used the term "debt," and from this point on, the topic is "guilt," concluding with the claim that the "rise of the Christian god as the maximum god that has been attained thus far," brings with it a "maximum of feelings of guilt into appearance on earth." (GM II:20)¹⁸

¹⁸ I recognize that the German word "Schuld" is means both "debt" and "guilt." Thus, in the original German there is no striking shift in this passage from "debt" to "guilt."

The abruptness and opaqueness of this significant development is stunning. Debt has been transformed, somehow, into guilt. And Nietzsche has left it to us to piece together an explanation. It is obvious that the transformation occurs through the influence of debt consciousness and religion. What is not as obvious is how debt consciousness, resentment, noble values, and giving form to the self interact with debt consciousness to produce consciousness of guilt.

Lastly, the nature of Nietzsche's presentation of the "moralization" of guilt and duty provides a crucial clue to the development and nature of the bad conscience. When Nietzsche presents the creditor-debtor relationship to the ancestors he describes the basic dynamic and then claims that the growth and success of the community alters the conception of the "ancestors." Initially the ancestors are thought of as present in spirit form. But as the community develops and is more successful they begin to conceive of the ancestral spirits as more and more powerful until they are transfigured into gods. (GM II:19) This escalating process continues, and in certain kinds of cultures – those that aspire to become universal empires, those that involve despotism – the monotheistic conception of god emerges. The culmination of this development is the "maximum God" of Christianity, which brings with it the "maximum of feelings of guilt." (GM II:20)

Nietzsche then presents a puzzle: It would not be unreasonable to assume that declining belief in God would involve a commensurate decline in feelings of guilt. But this is not the case. Nietzsche claims that this is because he has intentionally omitted a crucial feature in his explanation of the bad conscience: the actual moralization of the

concepts of guilt and duty. He describes this as, “their being pushed back into conscience, more precisely the entanglement of *bad* conscience with the concept of god.” (GM II:21) [original emphasis] The passage concludes with the Christian atonement as a solution to the problem of guilt, which Nietzsche claims was only temporary, obliquely suggesting that it only makes guilt worse. The upshot is that the moralization of guilt and duty makes it the case that guilt will not decline with loss of belief in God.

I will address this puzzle later in the chapter, but for now the important thing to notice is that Nietzsche’s presentation begins with primitive ancestral spirits and ends with the Christian conception of God (that than which no greater can be conceived). And he claims that he left moralization out of this account. So moralization must occur within the course of this development and must be a necessary condition for the development of the bad conscience. The indication that moralization occurs through the ascetic ideal is given in GM II:22, where Nietzsche claims that the bad conscience was “invented” by the human that suffers from the internalization of the instincts as a means to further self-torture. The interpretation that provides a pretext for self-torture is that the human being is guilty before God, in particular, because of the animal instincts, and this is clearly an ascetic interpretation of human existence.

2.5 The Development of the Bad Conscience

My general thesis¹⁹ on the development of the bad conscience is this: The slaves are under pressure to adapt to new conditions. On one front, they need to become obedient to new rules and customs in order to avoid punishment. As a result, internalized aggression is turned against the expression of any drives that would lead to disobedience. As they are increasingly successful they become more submissive, subservient, docile, and obedient, and adapted to their culture.

On the other front, the slaves are brought into the state with an already existing sense of debt consciousness. The creditor-debtor relationship has established a psychological connection between debt to the gods of a community and punishment. When the slaves inherit the concepts of “good and bad” and the nobles’ gods this puts them in a double-bind that produces resentment. Their efforts to adapt to the demands of the community have made them “bad” according to the value terms that they inherit, and this makes them incapable of paying their debt to the gods because they cannot live up to noble ideals. All of their efforts are “unworthy” because of their own inescapable “badness,” and the slaves cannot escape the pressure of debt consciousness and anxiety over punishment.

By using internalized aggression to dominate their other drives the slaves were able to satisfy the will to power through becoming obedient to customs. The underlying

¹⁹ I claim that Nietzsche offers us a general account of the bad conscience in GM II:16 & 17, but leaves a gap in his narrative between the mythical “blond beasts of prey” who create the first state, and the culmination of the development of guilt in the Christian doctrine of the atonement. I argue in Chapter Three that this gap is filled by the history of Exilic and Post-Exilic Judaism.

motivation was to avoid debt and punishment, and as far as their debt to the community, this was largely successful. But the slaves also feel debt to the gods and this is repaid through religious customs that include sacrifice and, more importantly, living up to the ideals of those gods. When the slaves accept the judgment that they are “bad,” they recognize that they cannot live up to noble ideals and so cannot repay their debt and avoid punishment. All of their efforts are in vain. This amounts to a systematic frustration of the will to power, which produces resentment.

Ressentiment seeks to identify the source of suffering and a target that the slaves can retaliate against. The ascetic priest provides the slaves with a target – themselves – by interpreting their suffering as God’s punishment for sin. Their feeling of impotence and worthlessness is the feeling of absolute guilt before God. They are powerless to avoid guilt and punishment, because of their essentially sinful animal natures. Thus, the animal bad conscience is interpreted into the bad conscience, and the idea of being guilty for sin provides the slaves with the pretext to engage in active self-cruelty (which explains why Nietzsche calls it the “active bad conscience”). The slaves turn the bad conscience toward the project of giving form to their own drives through negative self-critique and internalized cruelty against their own animal natures.

In what follows I will present these developments in more detail. Since, the frustration of slaves’ will-to-power and the development of resentment are produced in part by noble values, I will begin with a brief presentation of noble morality.

2.5.1 Noble Morality: “Good and Bad”

Nietzsche claims that the nobles coined the first value terms, which is the pre-moral distinction between “good and bad.” The development of these terms seems to occur well after the founding of the state. The creation of the state produces a stark class-division between the conquerors and the conquered, that will become the distinction between noble and slave. The immense difference between these groups – social, political, economic, ethnic, racial – evolves into a “pathos of distance” that is felt by the nobles and motivates them to mark the distinction between themselves and the slaves.

The pathos of distance clearly involves a recognition of the slaves’ “otherness” that provides the nobles with a contrast to themselves that stimulates self-awareness and is not merely descriptive. Nietzsche treats the pathos of distance as involving comparative self-worth that includes a feeling of elevation and a desire to be different than the object of comparison. It involves “looking down on subjects and instruments,” as well as “holding down” and “holding at a distance.” (BGE 257) The pathos of distance leads the nobles to recognize and affirm themselves as superior.

Initially, the nobles affirm themselves in terms of obvious social differences: they are the powerful, the lords, the commanders, the wealthy, and they affirm these differences as good. (GM I:5) Note that this is an entirely non-moral sense of “good.” Good simply indicates that the life of the nobles is to be preferred over the life of the slaves.

Over time, the feeling of social elevation is transformed into a feeling of “superiority of soul.” The nobles begin to think of themselves as superior, not simply in terms of

success and social status, but as the result of possessing certain character traits, or virtues (virtues in an *aretaic*, rather than moral sense). Their value judgments then pick out their typical character traits: they are the brave, the strong, the higher-ranking, the high-minded, the truthful, etc. This is where they begin to think of themselves as noble.

The nobles assessment of the slaves, according to Nietzsche, is secondary – an afterthought that simply picks out the opposite of what is noble and good, “in order to say ‘yes’ to itself still more gratefully and jubilantly.” (GM I:10) The slaves are seen as common, base, vulgar, lower-ranking, timid, and weak,. These judgments ultimately pass over into the concept “bad.” (GM I:4) Here again, the term “bad” is first used in a non-moral sense and simply indicates that the life of the slaves is not to be desired, and their traits are such that no one would choose to have them. But as the nobles’ value terms begin to reflect their sense of “superiority of soul,” they begin to think of the slaves as having inferior souls, and see them as low-minded, mendacious, and untrustworthy. At this point the judgment that the slaves are “bad” becomes an ethical judgment (though this is not yet a moral judgment, in Nietzsche’s sense).

2.5.2 The Source of Ressentiment

The bad conscience is a product of resentment, so the source and development of resentment must be explained in more detail. Nietzsche claims that bad conscience did not develop in the nobles, but that it would not have developed without them (GM II:17) The question is the way that the nobles produced the bad conscience.

A common view on the development of resentment is that it is produced straightforwardly by the heavy-handed rule of the nobles and the forced internalization of the instincts.²⁰ But Nietzsche claims that the transition into “society and peace” occurred so suddenly and inexorably that there was no struggle or resentment. (GM II:17) Transitioning into “society and peace” suggests that the nobles’ efforts to impose form on the slaves and enforce obedience was largely successful, thus I take it that this too occurred without struggle or resentment. This means that something else must be responsible for resentment. I claim that it is the dual influence of noble morality and noble religion.

A clarification is necessary before beginning. The claim that noble morality and religion are central to the development of resentment does not imply that the slaves needed to reflectively adopt and endorse the judgment that they were “bad,” or self-consciously “convert” to noble religion. It is sufficient, and more plausible, that they submissively mimicked noble religion to begin with (as Nietzsche claims), long before becoming conscious of the inheritance of debt. Similarly, it is sufficient that the slaves recognized that they were “bad” according to the nobles, which affected their standing in the community, their prospects for success, and their sense of comparative self-worth.

²⁰ This is clearly the case for Ridley. (Ridley, 2005, pp. 37-38)

Reginster’s account is more nuanced. On his account, the bad conscience results from the internalization of norms in addition to the internalization of the instincts, such that the person feels diminished self-worth over ethical failure. (Reginster, 2011, pp. 61-65) As far as I can tell, Reginster’s account does not factor resentment in as a source of the bad conscience, which is problematic, since Nietzsche makes it clear that the “man of resentment” has the invention of the bad conscience “on his conscience.”

Janaway, Conway, Risse, and May do not include resentment as part of their explanations for the bad conscience. fully explain the influence of resentment on the development of the bad conscience.

Nietzsche describes the slaves as predominantly “reactive,” compared to the “active” and creative nature of the nobles. They lacked their own value terms until the priests (more nobles) provide them with the concepts of “good and evil.” Thus, for a long period of time, the slaves simply lacked the means to conceive of themselves as in any way good. Nietzsche uses the concept of “inheritance” to describe their relation to noble values and their debt to the nobles’ gods. This does not require that the slaves actively adopt that inheritance, any more than one need adopt one’s own genetic inheritance. The slaves’ motivation is predominantly an instinctive adaptation to new conditions. I suggest that it be viewed analogously to the way that children inherit and adapt to the values and religious views of their parents. They absorb it unselfconsciously and find themselves thinking in the same terms without any reflective acceptance or understanding.

As we saw earlier, Nietzsche claims that when the slaves inherit noble values and adopt the nobles’ gods, the inheritance of unpaid debt to those gods overflows in all directions and marks the beginning of the feeling of guilt. (GM II:20) Nietzsche clearly equates the bad conscience with consciousness of guilt. (GM II:4) Since this passage claims that a feeling of guilt exists and grows, we can conclude that the bad conscience exists at this stage. And, since the bad conscience is a product of resentment, we can conclude that resentment has developed as well. The first task is to show how the value judgments “good and bad” and debt to the gods produced resentment through a systematic frustration of the slaves’ will to power. The second task is to show how resentment becomes part of debt consciousness and produces the bad conscience and

guilt. My claim is that this occurs when the slave find themselves trapped between the debt to the gods and a self-assessment that judges them unable to repay it.

As previously noted, it is relatively unsurprising that the slaves should inherit the concepts of “good and bad.” According to Nietzsche “good and bad” were the original value terms. The community that the slaves lived in was created by the nobles and structured according to their instincts for hierarchies of rank. To adapt to the conditions of this community it would be imperative for the slaves to recognize how “good and bad” were applied. As they became increasingly adapted to the nobles’ culture it would be surprising if they *did not* move from simply mimicking the use of these terms to accepting them as legitimate and true.

Moreover, recall that the transition into society and peace occurred without struggle or resentment. The nobles appear as an “inescapable doom,” and a “crushing and ruthless machinery” that goes to work forming the state. This suggests that the slaves were simply overwhelmed by the power, forcefulness, and energy of the nobles. Also recall Nietzsche’s claim that the vast majority of humans have an innate instinct to obey, compared to the very few with an instinct to command. (BGE 199) It is not implausible to imagine that the slaves instinctively regarded the nobles as they would any “alpha figure,” and were in fact impressed by, and in awe of the organization and accomplishments of the nobles. It seems natural that they would not dispute the right of the nobles to establish values and would then come to accept those values themselves, through the herd instinct to obedience.

Nietzsche also claims that the slaves inherit the psychological propensity for establishing orders of rank from the nobles. This is another indicator of the nobles' success in imposing form on the slaves, and it is unsurprising that the slaves should develop this propensity as well. Adapting to the hierarchical structure of the political state would require the slaves to learn that hierarchy, and the orders of rank established by the nobles, in order to navigate their new social environment. Obeying the nobles would require them to learn structures of authority, and carrying out their demands would require them to learn to prioritize certain matters over others. Accomplishing the tasks assigned to them by the nobles would likely involve establishing orders of rank within the slave community as well. Over time, learning to recognize and follow hierarchical structures would develop into the propensity to establish orders of rank themselves, as a way of adapting to their new conditions.

Of course, for the slaves, the main locus of this rank-establishing activity would lie in giving structure to their own drives and motivations. This connects establishing rank with a feeling of power and provides an even more compelling explanation for why the slaves inherit this propensity. As we have seen, subjugation forced the internalization of the aggressive drives and also put pressure on the slaves to adapt to new rules and customs. To do so they needed to control their own behavior by turning internalized aggression against the expression of the drives. In so doing they began to gain mastery over those drives, which established a hierarchy between them, and provided satisfaction to the will to power. Thus, the slaves have an incentive to engage in the activity of giving

themselves form, since it involves self-directed cruelty, which is the only means available to them to express aggression and satisfy the will-to-power.

Lastly, Nietzsche claims that the slaves adapted themselves to the gods of the nobles, and in so doing, inherited the pressure of unpaid debt. This too, seems like a natural consequence of the slaves' efforts to adapt to their new community and can be explained according to the logic of the creditor-debtor relationship, which is already a feature of their psychologies. According to this logic:

...every step toward the atrophying of the clan, all miserable chance occurrences, all signs of degeneration, of approaching dissolution always diminish the fear of the spirit of the founder and give an ever more reduced notion of his shrewdness, his foresightedness, and his presence as a power. (GM II:19)

Thus, the slaves would regard their own conquest and subjugation as strong evidence that their culture had degenerated and grown dissolute, and that their gods were weaker than the gods of the nobles. Fear of the powerful gods of the nobles, who are still present in the community that they now belong to, would motivate the slaves to conform to the religious customs of the nobles.

The logic of the creditor-debtor relationship to the gods has further implications for the slaves. According to that logic, the magnitude of the community's success is evidence for, and directly correlates with, the magnitude of the gods' power and is a measure of the community's indebtedness to the gods. Thus, the conquest of the slaves and the founding of the political state is evidence that the nobles' gods are very powerful. The fact that the nobles' gods have given them aid in founding the state means that the community owes a great debt to the gods. And when the slaves adopt these gods, they inherit this debt.

Nietzsche says that the slaves inherit the “pressure of this still unpaid debt and...longing for the redemption from the same.” This claim is clear from the reasoning provided above. The state is a great achievement and provides many benefits and protections to its members. The greater the benefits, the greater the debt, and the greater the fear that the gods have not been adequately compensated for these benefits. Thus, the slaves long to find a way to repay the gods and redeem themselves from this debt.

At this point, it is important to recall the nature of the creditor-debtor relationship to the gods and how it differs from the other forms of that relationship. Unlike the other forms, the relationship to the gods is a relationship to a creditor that is not present and does not actually communicate with the debtor. Thus, it is left to the imagination of the individual (an imagination of “growing fear,” Nietzsche claims (GM II:19)) to determine whether repayment has been adequate. Fear introduces the suspicion that it has not been adequate. Every success of the community serves to increase the fear that one owes still more to the gods. So, debt to the gods – even in the communities prior to the state – already involved anxiety over whether compensation had been provided.

Moreover, since the gods are understood to be the founders of the community and responsible for its continued success, the anxiety felt over debt is a collective one and places incredible pressure on the individual to conform to the customs of the community. Breaches of custom invite the displeasure of the gods and threaten the community as a whole. As I suggested earlier, this pressure creates a feeling of obligation that is nearly categorical. This only adds to the anxiety felt over providing proper repayment.

Lastly, the relationship to the gods promotes suffering as the essential means to repay the debt. Being obedient to customs was necessary to avoid the anger of the gods, but it was never sufficient. From the earliest stages the relationship to the ancestors required offerings and sacrifices in addition to obedience. As anxiety over debt and repayment increases, the magnitude of the sacrifices increases with it. A sacrifice is not genuine unless it involves loss and suffering. So the relationship to the gods involves the worry that one has not suffered enough to provide repayment and produces anxiety that more suffering will be required.

The problem is that the slaves cannot repay their debt to the gods.

The slaves have been caught in a double-bind from the beginning of their subjugation into the state. The pressure to adapt required them to take sides against their own dominant instincts and to control their expression. In the process of learning to obey new rules and customs they had to become submissive, subservient, and docile. These are exactly the traits that make them “bad” according to noble morality. And, having inherited “good and bad” from the nobles they are unable to see themselves as anything but “bad,” hence unable to repay the debt because they cannot live up to noble ideals.

Nietzsche claims that during the “middle period,” when the “noble clans take shape”, the nobles were able to repay their debts to the gods “with interest.” They do this, he claims, by returning “all of the qualities that had in the meantime become apparent in them, the noble qualities.” (GM II:20). This what Nietzsche refers to as the “ennobling” of the gods (which is not to be confused with their “hallowing”).

Nietzsche does not elaborate on the exact nature of this repayment, but an interpretation suggests itself. The nobles see themselves as “good,” “beautiful,” “happy,” and possessing “nobility of soul.” The claim that the repayment includes “interest,” suggests that they interpret their gods as having greater degrees of these qualities.

More intriguingly, Nietzsche claims that the nobles *returned* these qualities to the gods. This suggests that the nobles take the gods to have blessed them with these qualities in the first place. Nietzsche claims that even when the nobles seek out the opposites of their own qualities (in the slaves) they do so in order to affirm themselves “till more gratefully and jubilantly.” (GM I:9) So, returning these qualities to the gods seems to be motivated by self-affirmation, gratitude, and joy.

I want to suggest that this indicates that the nobles have experienced a changed relationship, and attitude, towards the gods and the debts that they owe to them. The nobles have come to see their debts, not as a burden to be repaid out of fear, but as indicators of their own strength and nobility, and as an opportunity to manifest their strength and gratitude through repaying their debts. Nietzsche claims that:

A people which still believes in itself still also has its own God. In him it venerates the conditions through which it has prospered, its virtues — it projects its joy in itself, its feeling of power on to a being, whom one can thank for them. He who is rich wants to bestow; a proud people needs a God in order to sacrifice Within the bound of such presuppositions religion is a form of gratitude. One is grateful for oneself: for that one needs a God. (A 16)

Thus, because of the qualities that the gods have bestowed upon them, the nobles are capable of performing deeds that enhance the community and are able to provide sacrifices that please the gods. If we recall that the “original conviction” of the

community is that it owes its existence to the “sacrifices and achievements of the ancestors” (GM II:19), then, in effect, the nobles repay the gods by contributing to their legacy. And because they are able to provide repayment “with interest” the nobles are able to relieve themselves of the pressure of debt consciousness.

This is not the case with the slaves. The slaves are bad according to noble values. If the means by which one repays the gods has some connection to possessing noble attributes and doing noble deeds that promote and advance the community and the legacy of the gods, then the slaves are hopelessly unable to repay their debt. The nobles, it seems, have gone through debt consciousness and have found a way to connect self-affirmation and the will to power with repaying the gods. The slaves, however, because of noble culture, are caught in a debt relationship that involves constant anxiety over their inability to repay debt through noble means, which will require them to repay that debt through punishment and suffering. (They, like the nobles, will go on to connect their debt to the gods with satisfaction of the will to power, but unlike the nobles, they will do so through self-negation, rather than self-affirmation.)

Thus, the slaves have been trapped between two conflicting sets of demands in noble culture. On one side, the pressure to conform to customs, in order to avoid debt to the community and punishment, forced them to become submissive, etc. through their own efforts to resist acting on their aggressive instincts. On the other side, these are precisely the slavish traits judged contemptible by the nobles. This leaves them incapable of acting in ways that would repay their debt to the gods. The slaves were forced to become “bad”

in order to fulfill their obligations to the community, but this leaves them unable to fulfill their obligation to the gods and unable to avoid debt through their own agency.

I take this situation to be the cause of resentment. The slaves already suffer from the internalization of the instincts and cruelty turned backward (the animal bad conscience). But they now also feel that repaying the gods is an obligation that they cannot fulfill, which produces a feeling of impotence, as well as anxiety over an unavoidable debt that carries the threat of punishment, all of which serves to increase suffering. The slaves are unable to locate a cause of their suffering to retaliate against and so cannot discharge resentment. They are unable to change their conditions and unable to change themselves. They are completely impotent to address their own suffering, or to find any meaning in it. Nietzsche claims that the problem is not suffering *per se*, but the meaninglessness of suffering, that:

...man, the bravest animal and the one most accustomed to suffering, does not negate suffering itself: he wants it, he even seeks it out, provided one shows him a meaning for it, a to-this-end of suffering. (GM III:28)

The complete inability to address suffering, or to retaliate against its cause, or to find a meaning and purpose in suffering, results in a systematic frustration of the slaves will-to-power, which produces resentment.

As resentment increases and no external outlets are found for its expression and discharge, it festers and poisons, producing the “human being of resentment,” who has the invention of the bad conscience “on his conscience.” We know that Nietzsche regards the bad conscience as synonymous with consciousness of guilt. We also know that it is

the ascetic priest who invents the idea of guilt. This indicates that, while the slaves do feel resentment, it is the resentment of the priest that is responsible for the concept of the bad conscience. The priest provides a meaning for suffering, and exploits the animal bad conscience, by interpreting it as punishment for sin, and guilt before God. The interpretation provided by the priest is what Nietzsche refers to as the “moralization” of guilt and duty, which entangles the bad conscience with the concept of God.

2.6 The Moralization of Guilt

As I mentioned in the section on the necessary conditions for the bad conscience, the moralization of guilt is important for understanding, not only the transformation of debt into guilt, but also for explaining why guilt does not necessarily diminish with atheism. The moralization of guilt also presents us with another puzzle. Nietzsche’s description of moralization suggests that the slaves had to have a conscience prior to the moralization of guilt and duty. Thus, they were sovereign individuals. If this is the case, it lends further support for my claim that noble values and religion produced resentment. It also has implications for how we understand Nietzsche’s presentation of the Greeks as a people who were able to “keep the bad conscience at arm’s length.” (GM II:23) Before addressing these issues, however, I will present Nietzsche’s claims about moralization.

Nietzsche describes the moralization of guilt and duty as “their being pushed back into conscience, more precisely, the entanglement of bad conscience with the concept of god.” (GM II:21) He also claims that guilt is felt toward God, and that this feeling grew in the same proportion as the concept of God, and human feelings concerning God. This

growth culminates with the concept of the maximum God – the God of Christianity – which brings with it a maximum in feelings of guilt. This suggests that guilt existed prior to the advent of the Christian God, and that the role of Christianity was to drive an already existing feeling of guilt to its greatest intensity.

Nietzsche describes the effects of moralization as follows:

...the prospect of conclusive redemption *shall* now pessimistically close itself off once and for all; the gaze *shall* now bleakly deflect off, deflect back from a brazen impossibility; those concepts of “guilt” and “duty” *shall* now turn themselves backwards — and against whom? There can be no doubt: first against the “debtor,” in whom bad conscience now fixes itself firmly, eats into him, spreads out, and grows like a polyp in every breadth and depth until finally, with the impossibility of discharging the debt, the impossibility of discharging penance is also conceived of, the idea that it cannot be paid off...finally, however, even against the “creditor,” think here of the *causa prima* of man, of the beginning of the human race, of its progenitor, who is now burdened with a curse (“Adam,” “Original Sin,” “unfreedom of the will”) or of nature, from whose womb man arises and into which the evil principle is now placed (“demonizing of nature”) or of existence generally, which is left as valueless in itself... (GM II:21)

This passage reflects the feeling of impotence in the face of inescapable debt that I claim produces resentment in the slaves. We also get a picture of the way the debt “overflows in all directions,” first affecting every aspect of the debtor, but then going further to be interpreted onto the “creditor,” i.e. the increasingly general “first causes” of human nature.

We also find specific references to Christian doctrines – “Adam,” “Original Sin,” – that are picked up by Nietzsche in the next aphorism, which is worth quoting at length, since it captures the effects of the bad conscience, as it occurs in the Judeo-Christian tradition, in explicit detail:

...that will to self-torment, that suppressed cruelty of the animal-human who had been made inward...who invented the bad conscience in order to cause himself pain after the more natural outlet for this desire to cause pain was blocked, — this man of bad conscience has taken over the religious presupposition in order to drive his self-torture to its most gruesome severity and sharpness. Guilt before God: this thought becomes an instrument of torture for him. In “God” he captures the most extreme opposites he can find to his actual and inescapable animal instincts themselves as guilt before God (as hostility, rebellion, insurrection against the “lord,” the “father,” the primal ancestor and the beginning of the world); he harnesses himself into the contradiction “God” and “devil”; he takes all the “no” that he says to himself, to nature, naturalness, the facticity of his being and casts it out of himself as a “yes,” as existing, corporeal, real, as God, as holiness of God, as judgeship of God, as executioner-ship of God, as beyond, as eternity, as torture without end, as hell, as immeasurability of punishment and guilt. This is a kind of madness of the will in psychic cruelty that has absolutely no equal: the will of man to find himself guilty and reprehensible to the point that it cannot be atoned for; his will to imagine himself punished without the possibility of the punishment ever becoming equivalent to the guilt; his will to infect and make poisonous the deepest ground of things with the problem of punishment and guilt in order to cut off the way out of this labyrinth of “*idees fixes*” once and for all; his will to erect an ideal — that of the “holy God” —in order, in the face of the same, to be tangibly certain of this absolute unworthiness. Oh, this insane sad beast man! What ideas occur to it, what anti-nature, what paroxysms of nonsense; what bestiality of idea immediately breaks forth when it is hindered only a little from being a beast of deed!... (GM II:22)

The main thing to note in this passage are the “anti-animal” and “anti-natural” themes that predominate, which are connected to self-loathing and punishment. These themes are evidence that the the bad conscience is connected to the influence of the ascetic ideal and that the bad conscience and the moralization of guilt is the product of the ascetic priest’s interpretation of an underlying phenomenon.

Since the moralization of guilt and duty is the work of the ascetic priest, I begin with a presentation of the priest and the ascetic ideal. Following that I will present the effects of moralization and address the two puzzles that are connected with moralization: Why

moralization makes guilt and duty immune to atheism, and how sovereign individuality is required to explain the difference between the Greeks and the slaves of GM II.

2.6.1 The Architect of Guilt: The Ascetic Priest

Nietzsche presents the nobles as divided into two distinct castes: the knightly-aristocratic caste (or warrior caste) and the priestly caste. The priests and warriors share the sense that they are superior, noble human beings and have the desire to rule over others. The warriors and priests also share the concept of noble morality in the sense of “good and bad.” Though with the priests the concepts of “good and bad” reflect their sense of “pure and impure,” rather than the warrior virtues of the knights.

In some cultures (Nietzsche frequently uses the example of the Hindu brahmins) the distinction between knights and priests resolves itself into hierarchy in which the priests rule. (There is reason to believe that Nietzsche might see this as preferable in many ways.) In other cultures there is tension between the warrior and the priestly castes, both of which desire to rule. The priestly caste is not suited to warfare, so it cannot engage in direct conflict with the warriors. (“Too bad for it when it comes to war!”) Out of the powerlessness of the priests “their hate grows into something enormous and uncanny, into something most spiritual and poisonous.” (GM I:7) Thus, the priests become “human beings of resentment” who seek revenge against the warriors. As the priests distance themselves further from the culture of the warriors, their resentment grows and they become hostile toward life. Nietzsche claims that:

...here a resentment without equal rules, that of an unsatiated instinct and power-will that would like to become lord not over something living but rather over life itself, over its deepest, strongest, most fundamental presuppositions... (GM III:11)

Through resentment, the priestly mode of evaluation eventually develops into the opposite of the warriors' values. "This process" Nietzsche claims, "is especially given an impetus every time the priestly caste and the warrior caste confront each other jealously and unable to agree on a price." (GM I:7) And it is through this process that the priest will eventually produce the ascetic ideal, which opposes everything that the warrior caste represents.

Nietzsche characterizes the ascetic ideal as world-negating, hostile towards life, not believing in the senses, de-sensualized,"(GM III:10), and claims that its three great "pomp words" are "poverty, humility, chastity." (GM III:8) We might sum it up as Clark does: the ascetic ideal is the idea that "the life of self-denial, the monkish life, is the highest human life." (Nietzsche, 1998, p. Introduction xxxi) The ascetic priest, is an "incarnate wish for a different existence, an existence somewhere else," (GM III:13) he:

...relates our life (together with that to which it belongs: "nature," "world," the entire sphere of becoming and transitoriness) to an entirely different kind of existence, which it opposes and excludes... (GM III:11)

The idea of an entirely different existence that is opposed to nature, the world, and the animal, has different conceptions (in the Rawlsian sense): Plato had one, Jesus had one, the Buddha had one, and even Kant had one. But the underlying concept is of a more fundamental reality, by which the natural, and human world, is judged faulty. This is the

ascetic ideal that is used by the ascetic priest to interpret the animal bad conscience (cruelty turned backward) to produce the guilty conscience.

2.6.2 The Ascetic Interpretation of the Animal Bad Conscience

As mentioned, the priests compete with the nobles for political power. Since they are unsuited for warfare, the priests need to employ other means. The slaves, and the interpretation of their situation through the lens of the ascetic ideal, are the priests' means to the end of political power.

The priests are sick like the slaves. Their will to power has been frustrated by the warrior caste and they suffer from the internalization of their dominant instinct (the will to rule over others) and feel from resentment. This makes the priests particularly well-suited to minister to the slaves. The priest, Nietzsche claims, is the "foreordained savior, shepherd, and advocate of the sick herd," whose realm is "dominion over ones who suffer." The priest must be:

related to the sick and short-changed from the ground up, in order to understand them – to get along with them, but he must also be strong, lord over himself more than over others, with his will-to-power intact, so that he has the confidence and the fear of the sick, so that for them he can be a foothold, resistance, support, compulsion, disciplinarian, tyrant, god. (GM III:15)

When the priest approaches the slaves he finds that part of his task is to defend it from itself, and from what "smolders within the herd":

...badness, deceitfulness, maliciousness and whatever else is characteristic of all the sick and invalids when among themselves; he fights shrewdly, hard, and secretly against the anarchy and ever-incipient self-dissolution within the herd, where that most dangerous blasting and explosive material, resentment, is constantly mounting and mounting. To discharge this explosive in such a way that

it does not blow up the herd or the shepherd, that is his true feat, also his supreme usefulness. (GM III:15)

What we find here is a picture of resentment in the slaves that has still not identified a source for its suffering – it has no target upon which it can vent anger, revenge, and punishment. Thus, resentment continues to grow, and threatens to result in indiscriminate destruction. Every sufferer, Nietzsche claims, including those who suffer from resentment:

...instinctively seeks a cause for his suffering; still more precisely, a perpetrator, still more specifically, a guilty perpetrator who is receptive to suffering — in short, some living thing on which, in response to some pretext or other, he can discharge his affects in deed or in effigy: for the discharge of affect is the sufferer's greatest attempt at relief, namely at anesthetization — his involuntarily craved narcotic against torment of any kind...one wishes, by means of a more vehement emotion of any kind, to anesthetize a tormenting secret pain that is becoming unbearable and, at least for the moment, to put it out of consciousness —for this one needs an affect, as wild an affect as possible and, for its excitation, the first best pretext. 'Someone must be to blame for the fact that I feel bad' — this kind of reasoning is characteristic of all those who are diseased, indeed the more the true cause of their feeling bad, the physiological one, remains concealed from them. (GM III:15)

As Maude Clark suggested in conversation, Nietzsche's view of human nature (especially prior to the moralization of guilt) seems to include the idea that human beings are not naturally inclined to blame themselves for their suffering, but rather, tend to look for a cause elsewhere.²¹ And what Nietzsche suggests in this passage is that human beings also tend to look for a cause of suffering that can be retaliated against and made to suffer in return.

²¹ Conversation with Maudemarie Clark, May 2018.

The problem for the slaves, in their specific context, is that there is no identifiable cause that meets that criterion. I have argued that the slaves have become obedient to customs, so the threat of punishment at the hands of the nobles is no longer a significant worry. Thus, the nobles cannot be identified as the cause of their suffering. I have also argued that the slaves suffer from resentment, because they feel themselves to be incapable of repaying an inescapable debt to the gods, which threatens them with punishment. But there is nothing that they can clearly identify as the gods' punishment. What the slaves suffer from is anxiety over the prospect of punishment, and anxiety, by its very nature, has no clearly identifiable source.

The genius of the priest is to identify a cause of the slaves' suffering that meets the criteria that resentment desires: a guilty perpetrator that is receptive to suffering; a living thing on which the slaves can discharge resentment and aggression. The priest does this by providing an interpretation of suffering that makes the slaves themselves the cause of their own suffering. To the claim: "I am suffering; for this someone must be to blame" the priest responds "That's right my sheep! Someone must be to blame for it: but you yourself are this someone, you alone are to blame for it — you alone are to blame for yourself!" (GM III:15)

The priest convinces the slaves that they are the cause of their own suffering, by interpreting their suffering as punishment for being guilty of sin. "Sin" is the "priestly interpretation of the animal's 'bad conscience' (cruelty turned backward)." (GM III:20)

The slaves suffer from a basic physiological condition – the internalization of the aggressive instincts – but the priest interprets all suffering into feelings of guilt, fear, and punishment. The pervasive anxiety the slaves feel over their inability to repay the gods is interpreted as the pervasive feeling of guilt over sin that deserves punishment.

The priest identifies animal nature as the source of sin and wrongdoing. The slaves are sinful because they have these instincts and they sin by acting on these instincts. Animal nature is interpreted as “guilt before God (as hostility, rebellion, insurrection against the ‘lord,’ the ‘father,’ the primal ancestors and beginning of the world.” (GM II: 22) The problem with the animal instincts, according to the priest, is that they lead human beings to value the natural world above the ascetic interpretation of reality, however that might be framed, though in this case: in terms of a transcendent and holy God. Acting on the aggressive instincts, or other base animal desires, is wrong, not because doing so tends to disrupt relationships within the community, but because it amounts to a failure to live up to God’s ideal of holiness. The animal instincts are the source of selfishness, lust, anger, violence, and greed, which run counter to the requirements placed on human beings by a holy God, who completely transcends animal nature, and the material order. “My kingdom is not of this world” summarizes the demands of an ascetic God.

By employing the ascetic ideal to interpret the slaves’ suffering as punishment for sin, the priest changes the direction of resentment. Once the slaves are convinced to see themselves as guilty before God they actively turn internalized aggression against their

animal instincts. This provides them with the pretext to engage in, and intensify self-cruelty, which satisfies the will to power.

Moreover, by interpreting suffering as punishment for sin, the priest gives a meaning and purpose to the slaves' suffering. Nietzsche claims that the ascetic ideal "saved the will," by providing human beings with a meaning to existence, that gave meaning to their suffering. It was the meaninglessness of suffering, Nietzsche claims, that was the "curse" on humanity, and "the ascetic ideal offered it a meaning." The human being was no longer "like a leaf in the wind, a plaything of nonsense," but had a sense of purpose, and a goal: to deny animal nature in order to become more "holy."

Though, of course, Nietzsche warns us not to conceal from ourselves what the ascetic ideal expresses:

...hatred of the human, still more of the animal, still more of the material, this abhorrence of the senses, of reason itself, this fear of happiness and of beauty, this longing away from all appearance, change, becoming, death, wish, longing itself — all of this means...a will to nothingness, an aversion to life, a rebellion against the most fundamental presuppositions of life...(GM III:28)

It is also important to recognize what the ascetic interpretation of existence has produced in human psychology. The bad conscience, as a guilty conscience, is an interpretation of an underlying phenomenon. The slaves felt trapped in a self-reinforcing dynamic of impotence, debt, badness, resentment, and suffering. The bad conscience is an interpretation of this basic feeling of personal inadequacy, and helplessness in the face of meaningless suffering, as punishment for sin. This provides a meaning for suffering, but it reinforces and preserves the all-encompassing feeling of worthlessness and impotence.

This is why Nietzsche claims that the priests medication is a “mere affect-medication, it cannot be a matter of a true healing of the sick in the physiological sense,” (GM III:16) and that the priest “combats only suffering itself, the listlessness of suffering, not its cause, not the actual state of sickness.” (GM III:17)

Through the priestly interpretation the human being remains impotent to please God and repay the debt, because the human being is a sinner. Because of sin, punishment is still without measure. There is no hope that punishment will ever be equal to one’s guilt. Thus, the concept of sin functions primarily as a pretext for self-punishment that satisfies the will to power and anesthetizes suffering through violent affect. Through the concept of sin, human beings are provided with an inexhaustible opportunity to engage in self-cruelty, because they can never live up to the ascetic ideal and so will always feel guilt. Thus the guilty conscience just is the psychological connection between the sense of culpable worthlessness that deserves punishment, which only serves to mask the fact that it is entirely in the service of self-cruelty.

2.7 The Atonement as Temporary Remedy

GM II:21 first presents the moralization of guilt and then concludes with a brief mention of the Christian doctrine of the atonement. The doctrine of the atonement appears after Nietzsche presents the escalating effects of moralization, which first turns guilt against the debtor/sinner, but continues by assigning guilt to the various “creditors” of human existence: the primal ancestors and progenitors of human nature (“Adam and

Eve”), the natural world, and finally existence itself, which is judged valueless. Nietzsche then presents the atonement as the:

...paradoxical and horrifying remedy in which tortured humanity found temporary relief, Christianity’s stroke of genius: God sacrificing himself for the guilt of man, God himself exacting payment of himself, God as the only one who can redeem from man what has become irredeemable for man himself — he creditor sacrificing himself for his debtor, out of *love* (is that credible?—), out of love for his debtor!...(GM II:21)

It seems clear that the doctrine of the atonement is connected with the process of moralization, but Nietzsche does not make the connection explicit, leaving it once again for the reader to discern the connection from comments made in other passages. What Nietzsche does say is that the remedy provided by the atonement is “paradoxical and horrifying” and that it is only temporary, which means that it does not put an end to guilt, and as I will argue, does not put an end to punishment. The task is to determine why this is so, and the implications this has for the moralization of guilt.

We can begin by noting the continuing influence of the creditor-debtor relationship. In this case, God is the creditor who advanced good to humanity (every good...), and human beings owe God repayment in the form of obedience to the ideal of sinlessness and holiness. Human beings, however, fell into sin. Sin is a violation of the contract with God, so it must be punished. But the debt that is owed to a maximum God is maximal itself, and having corrupted their natures through sin, human beings are no longer capable of being holy and sinless, so the debt is impossible to repay. Human beings not only lack sufficient funds to repay the debt, they also lack the appropriate currency. The debt cannot be repaid through living up to God’s ideal and it cannot be repaid through

sufficient suffering, since the debt is infinite.²² But despite the fact that human beings deserve infinite punishment and suffering for sin, God loved human beings enough to take on the punishment for sin.

The doctrine of atonement is clearly meant to address the slaves' longing for redemption from unpaid debt (GM II:20), and resonates with the claim that debt consciousness periodically led human beings to desire a wholesale redemption through some enormous counter-payment. (GM II:19) Of course, with the atonement, human beings see themselves as utterly inadequate to provide repayment, leaving that to God. Thus, doctrine of the atonement reflects the earlier community's anxiety over providing adequate repayment to the gods, turned pathological in the political state, through the slave's resentment, internalized aggression, and the need "to be tangibly certain of his absolute unworthiness."²³

²² Conway notes that the Christian God is maximal not only in the sense of what is demanded from human beings, but also in the sense of what is not accepted. (Conway, 2008, p. 142)

²³ That the doctrine of atonement primarily appeals to the need to feel utterly unworthy is reflected in the fact that the sinner's debt is fully paid by God. Conway points out that the same sinners who claim to assume maximum responsibility for their wrongdoing actually make themselves radically irresponsible, because they categorically refuse to take on the burden of any debt that they might be expected to repay. Sin, he claims, keeps the burden constant and unchanging, so it does not threaten to expose the sinner to any new and potentially overwhelming debt. Thus, guilt is a means to protect the self from the possibility that some responsibility might be too much to bear. (Conway, 2008, p. 143)

There is much in this claim that is insightful, though I worry that that it overlooks differences in Christian theologies, some of which include emphasis on the sinner's responsibility to work out their salvation in various ways. In Roman Catholic theology, sanctification (becoming righteous and holy) precedes justification (God removing the sinner's guilt and rescinding the penalty for sin), which entails that the sinner must take on the responsibility for works of righteousness through their own effort. While most Protestant theologies maintain that justification precedes sanctification, many also emphasize the need for the sinner to be responsive to the workings of the Holy Spirit and to cooperate with its efforts to lead the sinner into sanctification and holiness. Thus, while the sinner may be justified completely through faith and grace, spiritual progress is, at least in part, a matter of the sinner's own efforts and is their responsibility.

The Christian accepts the idea that God loves humankind and would promote the idea that the atonement is God's remedy for the human predicament. If the problem is punishment for sin, then God provides a solution by taking on the punishment and forgiving human beings for their sins. One might well imagine the relief the repentant sinner feels when the sentence of eternal punishment is lifted.

One might also think that forgiveness would spell an end to guilt over sin. But, again, Nietzsche claims that this remedy was only temporary, which means that it did not put an end to guilt (or, as we will see, punishment).

There are several problems with the doctrine of the atonement as a cure for guilt. One obvious problem is the claim that God took on the punishment for sin, which leaves guilt for sin unaddressed. This is so, even if a particular theology claims that God also took on the guilt for sin, because the doctrine of the atonement necessarily entails human guilt.

That the doctrine of atonement is explicitly only about punishment should be fairly obvious. The claim is that human beings not only sin, but are tainted with the curse of original sin through their own free choice, and are no longer able to live up to God's ideal through the efforts of their own will. The punishment for sin is eternal damnation. Repentant human beings are saved when God takes on the punishment for sin.

The fact that human beings need to be "saved" from punishment indicates that they *are guilty* and deserving of punishment. Commuting a sentence does not address the guilt of the agent; at most it shows mercy. Forgiveness of sins, by the very definition of forgiveness, requires the judgment that the one forgiven was guilty of sin. Forgiveness of

sins does not remove the fact of guilt, it only shows willingness to forego the right to justified retaliation. Moreover, in orthodox Christian doctrine, even a redeemed, sanctified, and “perfected” Christian, who through God’s grace no longer commits sinful acts, is still tainted by the fact of sin, which nothing can ever fully erase. (See Kant’s *Religion Within the Bounds of Mere Reason*, for a more secular analogue to this point.) Thus, taking on the punishment for guilt does nothing to address the fact of guilt.

Eliminating the threat of divine punishment, through the atonement, is also irrelevant in an important sense. It does not address the more pervasive form of punishment that remains. When the slaves accept the interpretation of their suffering as guilt before God, they thereby accept the idea that their sinful animal natures are the source of sin and are deserving of punishment. This provides the pretext to satisfy cruelty by engaging in self-punishment. Eliminating the threat of divine punishment does nothing to eliminate the more pervasive form of punishment which is internalized in the human conscience.

Moreover, divine punishment is an external threat and, as such, would work against the priest’s efforts to redirect resentment in several ways. First, the threat of divine punishment addressed by the atonement is eternal damnation in the afterlife, which can never be made manifest in this life. In this case, punishment remains a looming threat that only serves to increase anxiety and foster resentment. And since God is the source of

the threat, resentment would be directed toward God²⁴, rather than back towards the self. But God cannot be retaliated against and made to suffer, so resentment cannot be satisfied and would continue to grow. The priest needs resentment to be directed against the self, and not against God, so the priest needs to offer deliverance from eternal punishment.

In addition, as an external threat, divine punishment would not provide the slaves with any pretext to internalize their own aggression. If eternal damnation for sin is unavoidable, then there is nothing for the slaves to do about it, and so no reason to direct aggression within. The slaves would simply continue suffering, their resentment would lack any viable direction, and this would either erupt in anarchic violence, or produce listlessness and depression (perhaps alternating bouts of both). To avoid this, the priest needs guilt and punishment to be internalized and so needs to eliminate the threat of eternal punishment.

Finally, the real problem for human beings is not punishment and suffering. Nietzsche claims that the human does not negate suffering in itself, but wants it and even seeks it out, provided it has a meaning. (GM III:28) So the problem is the meaning of suffering; and guilt is the interpretation that has provided the slaves' suffering with meaning. So the

²⁴ Consider Luther's thoughts on the anger and hatred he felt toward God the Father, who demanded perfect obedience and threatened eternal punishment for failure. "Though I lived as a monk without reproach, I felt that I was a sinner before God with an extremely disturbed conscience. I could not believe that he was placated by my satisfaction. I did not love, yes, I hated the righteous God who punishes sinners, and secretly, if not blasphemously, certainly murmuring greatly, I was angry with God and said, "As if, indeed, it is not enough, that miserable sinners, eternally lost through original sin, are crushed by every kind of calamity by the law of the decalogue, without having God add pain to pain by the gospel and also by the gospel threatening us with his righteousness and wrath!" (Luther, 1962, p. 11)

doctrine of the atonement cannot eliminate guilt, because it relies on the interpretation of suffering as guilt, to provide meaning to the slaves' suffering. This interpretation is largely successful because redirecting resentment and aggression against the self satisfies the will to power and provides a goal for the human will. Thus, the doctrine of the atonement cannot be designed to eliminate guilt. Guilt is required to give meaning to suffering and as the pretext that justifies internalized aggression against the self. The doctrine of the atonement can only address punishment, but even in this, as we have seen, it is only an irrelevant source of punishment that is addressed. This is part of the priest's genius.

Thus, the doctrine of the atonement seems designed to only provide temporary relief from either guilt or punishment. Guilt is never actually addressed, any sense of relief from guilt that is felt is only the result of relief from the threat of punishment. But this sense of relief is only an anesthetic that serves to temporarily mask the feelings of guilt. Punishment cannot actually remove the fact of guilt, so it cannot actually remove the feeling of guilt. Once the anesthetizing effect wanes, the recognition that one is guilty remains. Similarly, once the external threat of divine punishment is eliminated, this allows the dynamic between guilt and punishment to be fully internalized and made instinctive, thus preserving the most pervasive sense of punishment in the human conscience.

The Christian might respond to the claim that the atonement preserves the concepts of guilt and punishment, by directing attention to God's motivation: God engages in self-

sacrifice out of love for humanity. This is one of the central claims of the doctrine of the atonement. To ignore this facet of the doctrine is a failure to see how love cancels out guilt by taking on punishment, offering forgiveness, and restoring human beings' relationship to God. The God who loves humanity this much is not concerned with the guilt, so human beings should be freed from it as well.

This introduces an especially obscure feature of Nietzsche's account of the atonement. He appears especially baffled by the idea that a creditor would take on punishment *out of love* for a debtor. He follows up on this when he says:

... whoever is still capable of hearing (but one no longer has the ears for it today! —) how in this night of torture and absurdity the cry love resounded, the cry of the most longing delight, of redemption in love, will turn away, seized by an invincible horror... There is so much in man that is horrifying! ... The earth has been a madhouse for too long! ... (GM II:22)

Why does Nietzsche find the idea that the creditor might love the debtor enough to take on punishment for debt so baffling, and indeed horrifying? I suggest this answer: The idea of a "creditor's" love for a "debtor" raises concerns as to the legitimacy of calling this a creditor-debtor relationship in the first place. One might think that, if the creditor genuinely loved the debtor to begin with (or, I suppose, came to love the debtor along the way), they would not be in a creditor-debtor relationship. They would be "lover" and "beloved." And whatever goods the lover possesses would be freely given to the beloved. Love makes the goods a gift, not a loan, so there would be no debt owed.

A genuine gift, it seems, should not require repayment in kind. One might think that gratitude is owed as a response to a genuinely loving gift and that an ungrateful recipient

is guilty of some kind of failure. But, even if this is the case it is implausible to think that a person genuinely motivated by love, without ulterior motives, would seek punishment for ingratitude, even if one felt that gratitude was “owed” in some sense. The thought that punishment is appropriate for this failure would reveal that the relationship was really between a would-be creditor and their intended debtor, and was not a relationship of love. So, it is either a relationship of lover to beloved, or creditor to debtor, and it cannot be both.²⁵

The doctrine of the atonement conflates these two relationships. It places God in the role of both lover of humanity and humanity’s creditor. In the tension between these two roles it is clear that love is not sufficient to cancel the debt. God’s own nature is understood to demand that guilt be punished one way or another. If human beings are not to be punished for their guilt, then God must take on the punishment by punishing God.

This must be part of what is both paradoxical and horrifying to Nietzsche. Forgiveness and love, it seems, should dictate that punishment is not required at all. But God “forgives” and then engages in self-punishment, which makes it seem that it is not really forgiveness at all. God’s anger against sin must be satisfied and love is in fact *not sufficient* to cover a multitude of wrongs. Only punishment can do that. Thus, the doctrine of atonement provides a model for human beings to engage in self-inflicted punishment over guilt. This effectively enshrines, insulates, and even deifies the primitive

²⁵ This idea is echoed in another passage where Nietzsche says “What? A god who loves men provided that they believe in him and who casts evil gazes and threats at anyone who does not believe in this love? What? A love hemmed in by conditions as the feeling of an almighty god? A love that has not even mastered the feeling of honor and roused vengefulness? How oriental this all is! ‘If I love you, what does that concern you?’ is surely a sufficient critique of all of Christianity.” (GS 141)

thought that there is always an equivalence between debt and punishment and that every debt must be paid off. Not even God can escape the law of punishment for guilt. Thus, punishment is divine.

It was precisely this need for a connection between punishment and guilt that produced the “symbol of the ‘holy cross’, that gruesome paradox of a ‘god on the cross,’ that mystery of an inconceivable, final, extreme cruelty and self-crucifixion of God for the salvation of man.”²⁶ (GM I:8) The doctrine of the atonement is itself the culmination of the desire for self-cruelty that needs an unbreakable connection between punishment and guilt to ensure its continuation. The atonement simultaneously drives guilt to its greatest extreme (what greater crime is there than killing God?), and by eliminating the threat of divine punishment for guilt, places the power of punishment permanently in the hands of the slave of bad conscience.

2.8 The Implications of Moralization: Guilt Survives Atheism

We are now in a position to explain a puzzle about the moralization of guilt. Nietzsche describes the moralization of guilt and duty as “their being pushed back into conscience, more precisely, the entanglement of bad conscience with the concept of god.” (GM II:21) And as we have seen, the moralization of guilt is accomplished through the interpretation provided by the ascetic ideal. Nietzsche clearly thinks that the ascetic ideal is a widespread phenomenon on earth. He claims of the ascetic ideal that:

²⁶ Thanks to Gavin Lawrence for highlighting the idea that the atonement is the self-crucifixion of human nature, by suggesting that his students imagine how a visitor from another planet would react, upon finding a church in nearly every village, town, or city in the Western world that has as its centerpiece, the agonized form of the Earth’s dominant species hanging from an instrument of torture.

Such a monstrous manner of valuation is not inscribed into the history of humankind as an exception and curiosity: it is one of the broadest and longest facts there is. Read from a distant star the majuscule script of our earthly existence would perhaps tempt one to conclude that the earth is the true ascetic star, a nook of discontented, arrogant, repulsive creatures who could not get rid of a deep displeasure with themselves, with the earth, with life and who caused themselves as much pain as possible out of pleasure in causing pain: — probably their only pleasure. (GM III:11)

There are different conceptions of the ascetic ideal, not all of them related to the Judeo-Christian religious tradition. Nevertheless, in his presentation of the moralization of guilt, Nietzsche focusses on the Judeo-Christian tradition, where the form given to the ascetic ideal is a transcendent and holy God.

The puzzle that emerges from the text is Nietzsche's claim that the feeling of guilt does not decrease as belief in the Christian God declines, as one might plausibly think it would. After all, if guilt is "guilt before God," and God is no longer there to judge one guilty, it seems that the feeling of guilt should wane and disappear. But Nietzsche claims that the moralization of guilt prevents this from happening.

We can begin by examining the ambiguity of the claims that guilt and duty are "pushed back into conscience." "Pushed back into conscience" could mean that guilt and duty had somehow become foreign to the conscience, and moralization reversed this development. But since the slaves already had an instinct to obedience prior to their subjugation into the state, and since one of their main concerns was to adapt and become obedient to the rules of the state, it hardly seems plausible that the idea of obligation or duty had become less than a central concern.

The more plausible reading is that “pushed back into conscience” means that the concepts of guilt and duty were already present in the conscience, but moralization made them central – they were pushed further, and deeper into the consciences of the slaves, and became a pervasive, and instinctive feature of their thinking. According to the account I have given, human beings who were subjugated into the state already had an instinct to obedience – a primitive, non-moral, feeling of obligation to follow the rules of the community – and were capable of doing so. In the context of the community, the rules and expectations on one’s behavior are fairly clear and explicit, and once the individual was acculturated, following the rules became fairly instinctive.

But with new rules and customs and the internalization of the instincts, this process begins again in a new register, as the human animal begins to rely on conscious thinking to avoid any expression of the instincts that might lead to disobedience. This makes the behavior of the earliest members of the state much less instinctive and much more present to conscious thought. To adapt completely to the conditions of society and peace, it is necessary for obedience to customs to become instinctive once again. But this is complicated by the internalization of the aggressive instincts. Note the obvious fact here: aggression is instinctive in these human beings, but no longer has an acceptable outlet for expression. Thus, the internalization of aggression must be made instinctive as well, for the individual to adapt to society.

Ressentiment, however, complicates the task of making internalized aggression instinctive. The slaves already suffer from internalized aggression and the animal bad

conscience. The influence of noble morality and religion produces anxiety over punishment, because they perceive themselves as inadequate to pay their debts to the gods. This produces resentment. Resentment instinctively seeks a cause of suffering against which it can retaliate and be discharged. But the slaves lack an identifiable target, so their resentment results in even more internalized aggression and produces even more suffering. Lacking a target, resentment threatens to erupt in any and all directions, on the basis of any available pretext.

The priests successfully prevent this by redirecting resentment. They accomplish this by convincing the slaves that they are the culpable source of their own suffering and deserving of punishment. They give the slaves a way of seeing themselves as “wrong,” which provides a pretext for self-inflicted cruelty. By providing this interpretation the priests give internalized cruelty a meaning and purpose, and thus a “justification.” Redirected in this way, internalized cruelty is made instinctive in the human animal, i.e., it is pushed deeper into the conscience as one of its central features.

The fact that self-inflicted cruelty is made instinctive is nearly sufficient to explain why Nietzsche believes that guilt endures after theistic belief declines. But there must be more to it than mere instinct. As we saw with the slaves, merely internalized cruelty was a problem, since it led to suffering without meaning. Thus, it seems that it is part of human nature, or perhaps, part of sovereign individuality, to need a pretext – a reason – to engage in self-directed aggression. Thus, something that provides a pretext for self-directed aggression must survive the death of God.

I take this pretext to be found in the fact that the priest interprets animal nature as the source of sin. Once this idea is accepted, it can be understood in increasingly general terms, such as selfishness, egoism, or even Kant's "self love," and these become the source of moral wrongdoing – none of which need have any connection to God or sin. When selfishness is identified as the source of moral wrongdoing, and selfishness is understood as connected to our base animal drives, then the human being can continue to feel pervasive guilt for being inadequate to deny selfishness and act on moral principles. So what is essential to moralization is not its connection to God. God was a contingent feature of a particular cultural interpretation of the ascetic ideal. What is essential to moralization is that a sense of dissatisfaction over human nature, and a feeling of being impotent to overcome one's own nature, has been made a central feature of the conscience and guilt. The feeling of dissatisfaction and impotence can be interpreted in ways that do not involve God, thus, even without God, the human being may continue to feel pervasive guilt, and this allows the human being to continue to engage in ever more sophisticated and subtle acts of self-punishment and self-cruelty.

2.9 The Bad Conscience: A Sickness Like Pregnancy

While troubling in the extreme, Nietzsche clearly thinks that the bad conscience holds some promise for human beings. "It is a sickness, bad conscience — this admits of no doubt — but a sickness as pregnancy is a sickness." (GM II:19) To claim that the bad conscience is like pregnancy suggests that, while the gestation period of the bad conscience might be replete with pain, discomfort, nausea, etc., it has the potential to

produce something new and hopeful, perhaps even beautiful. But we should be clear about the nature of the sickness first.

The bad conscience could be described as a kind of feedback loop that covertly strengthens what it is meant to undermine. Initially, it seems, the aim of the slaves is to gain control over the expression of aggression, since it leads to violations of the rules, debt, and punishment. This itself involves the internalized use of aggression. But with the invention of the bad conscience, aggression itself is seen as a source of debt, or guilt, and internalized aggression is turned towards the denial and repudiation of the aggressive instincts. This secretly satisfies and strengthens the aggressive instincts and the desire to make suffer and gives the person an incentive to engage in this particular project of “self-improvement,” which is in actuality a project of self-cruelty.

Since the aggressive instincts themselves are consciously rejected as a source of debt, or guilt, this activity, monitored through self-reflection and increasing self-awareness, becomes more and more covert and subtle. Cruelty is given increasingly sophisticated disguises. Its greatest disguise is altruistic morality. Nietzsche claims, “...we know one thing henceforth...namely what kind of pleasure it is that the selfless, the self-denying, the self-sacrificing feel from the very start: this pleasure belongs to cruelty.” (GM II:18) The unintended, and ironic, result of this project of “moral self-improvement” is that the aggressive instincts remain in an unacknowledged position of dominance.

Nietzsche gives a few hints as to the potentially positive outcome of the bad conscious. Even as he details the self-cruelty involved in the slaves' project of giving form to themselves, he points out that the active bad conscience:

As the true womb of ideal and imaginative events, finally brought to light... a wealth of new disconcerting beauty and affirmation and perhaps for the first time beauty itself ... For what would be "beautiful" if contradiction had not first come to a consciousness of itself, if the ugly had not first said to itself "I am ugly"? (GM II:18)

This suggests that, even at its worst, the bad conscience is responsible for creating concepts that we take to be important (Nietzsche certainly takes beauty to be important). The negativity, critique, and even contempt associated with the bad conscience is capable of producing ideals that are worth living by.

This same theme is brought up when, immediately after pointing out that the bad conscience amounts to a "declaration of war against the old instincts," Nietzsche notes that:

...on the other hand, with the appearance on earth of an animal soul turned against itself, taking sides against itself, something so new, deep, unheard of, enigmatic, contradictory, and full of future had come into being that the appearance of the earth was thereby essentially changed. Indeed, divine spectators were necessary to appreciate the spectacle that thus began and whose end is still by no means in sight — a spectacle too refined, too wonderful, too paradoxical to be permitted to play itself out senselessly-unnoticed on some ridiculous star! Since that time man is included among the most unexpected and exciting lucky throws in the game played by the "big child" of Heraclitus, whether called Zeus or chance — he awakens for himself an interest, an anticipation, a hope, almost a certainty, as if with him something were announcing itself, something preparing itself, as if man were not a goal but only a path, an incident, a bridge, a great promise... (GM II: 16)

While certainly a bit of poetic indulgence, this passage clearly indicates that Nietzsche takes the bad conscience to hold promise. An “animal soul turned against itself” is capable of wanting to change its own nature, and although this desire can be motivated by self-undermining resentment, there is nothing that precludes a different, more productive, motivation. The mention of the gods suggests the creation of ideals that give meaning to, and can inspire, human development. Most importantly, it seems, the bad conscience is responsible for human beings seeing themselves as having a future that involves advancements in human nature, rather than being an evolutionary dead-end. The bad conscience prevents the self-satisfaction, complacency, and stagnation that is characteristic of “the last-men” of *Thus Spake Zarathustra*.

The central problem with the bad conscience is that it developed out of animal suffering and was then directed at our animal natures, interpreted as the source of guilt and suffering.

We modern humans, we are the heirs of millennia of conscience-vivisection and cruelty to the animal-self: in this we have our longest practice, our artistry perhaps, in any case our sophistication, our over-refinement of taste. For all too long man has regarded his natural inclinations with an “evil eye,” so that in him they have finally become wedded to “bad conscience.” (GM II:24)

I would not go so far as to claim that this is the only problem with the bad conscience. Any psychological dynamic that exploits an insoluble situation, targets an intractable feature of the self, and does so as a means to sustain a self-perpetuating project of self-cruelty (or cruelty to others for that matter) is an unhealthy phenomenon, even if the target was not animal nature. But it is certainly clear that our antipathy to our animal

natures is the worst feature of the bad conscience, and should be eliminated, if possible.

Nietzsche claims that it is, but it will be difficult:

A reverse attempt would in itself be possible — but who is strong enough for it? — namely to wed to bad conscience the unnatural inclinations, all those aspirations to the beyond, to that which is contrary to the animal — in short the previous ideals which are all ideals hostile to life, ideals of those who libel the world. (GM II:24)

Interestingly enough, this does not involve the elimination of cruelty, but rather, its sublimation towards achieving healthier ideals:

For this goal one would need a different kind of spirits than are probable in this of all ages: spirits strengthened by wars and victories, for whom conquering, adventure, danger, pain have even become a need...for this one would need a kind of sublime malice itself, an ultimate most self-assured mischievousness of knowledge, which belongs to great health; one would need, in brief and gravely enough, precisely this great health! (GM II:24)

Chapter 3: Historical Instances: Post-Exilic Judaism and the Ancient Greeks

“Then devise for me the love that bears not only all punishment but also all guilt! Then devise for me the justice that acquits everyone except the one who judges!”²⁷

Introduction

A key hermeneutical principle for interpreting Nietzsche is that it is often the case that what he omits from an account is as important for understanding him as what he includes, and that the omission is intentionally left for the reader to bring to light. With this in mind we should note a peculiarity in Nietzsche’s account of the development of the bad conscience which is the historical gap between his thoughts on the origins of the bad conscience and its culmination in Christianity and the doctrine of the atonement.

Nietzsche’s presentation of the development of the bad conscience is rather general and historically unspecific. We are told that it is obvious (“it goes without saying”) that the state was created by “some pack of blonde beasts of prey, a race of conquerors and lords (GM II:17), who conquered some nomadic tribal groups, which may have been older “cattle raising” or “trading” cultures. (BGE 257) We get an account of the internalization of the instincts and a description of the bad conscience at work. In a very compressed account of religious developments through the creditor-debtor relationship to the ancestors, we quickly arrive at the Christian doctrine of the atonement and all of the madness it produced. So, somehow, the primeval (and nearly mythical) events that involve the “blonde beasts” lead up to events that occur in the early days of the First

²⁷ Z:I:19

Millennium CE. And somewhere in the time between the primeval past and the time of Christ, the slave revolt in morality occurred and the bad conscience was interpreted as guilt before God. The issue is whether we can pick up the trail of this development and point to specific historical developments that produced the bad conscience and can be traced to its culmination in Christianity.

I want to suggest that, without telling us so explicitly, Nietzsche has presented us with a general formula for the internalization of the instincts and the development of the animal bad conscience in GM II:16 & 17 (and perhaps aspects of GM II:18), where we find the “blonde beasts” and the conquered nomads. This general formula (along with the religious presuppositions presented in GM II:19 & 20) is meant to guide our understanding of the specific (though incomplete) historical account that he presents in GM II:21 & 22, where we find the moralization of guilt and explicitly Jewish and Christian concepts. The implication of this reading is that internalization of the instincts need not be a one-time event and is not exclusively limited to the basic forms of the primitive instincts that humans share with other animals, such as aggression, but can occur when the primitive drives have already been incorporated into a hierarchical structure and intertwined with other drives.

According to Nietzsche’s drive psychology, basic drives can form “alliances” with other basic drives. (I have argued that the instinct to obedience and the instinct for freedom form this kind of alliance in the slaves.) When these alliances successfully satisfy the will-to-power, the alliance between basic drives can become a stable and

instinctive feature of the person's overall drive economy, effectively producing a new drive (such as the conscience). Basic drives can be sublimated into "coalitions" of drives that seek to fulfill some "higher" function of the person, such as the drive to knowledge, or value-drive.²⁸ On my reading, these "coalitions" of drives can also be internalized when the activities they are in service of, are denied expression and forced to find other ways of being satisfied. This also implies that internalization can take place repeatedly for a person, or a community of people, as they are subjected to different forms of constraint on their activities and dominant instincts.

3.1 Post-Exilic Judaism and the Bad Conscience²⁹

Nietzsche claims that the "slave revolt in morality" began with the Jews and has a two-thousand year history behind it. This mention of a two-thousand year history narrows the possible historical options that he might have in mind.

The Hebrew Bible gives an account of a kind of slave revolt in Exodus, where Moses led the Hebrew slaves out of bondage in Egypt. It seems unlikely that Nietzsche has this in mind for several reasons. First, the only detailed account of the Hebrew exodus is

²⁸ Nietzsche describes the philosopher's drives as including his "doubting drive, his negating drive, his wait-and-see ('ephectic') drive, his analytical drive, his exploring, searching, venturing drive, his comparing, balancing drive, his will to neutrality and objectivity, his will to every '*sine ira et studio*' [without anger and partiality]. Here we get a picture of the many drives that are incorporated into, and involved in the "single" "higher" activity of doing philosophy.

²⁹ The support for this interpretation is provided by Avery Snelson's article "The History, Origin, and Meaning of Nietzsche's Slave Revolt in Morality." Snelson argues that the slave revolt should be understood as occurring in two stages, beginning with the prophetic interpretation of Israel's political situation (primarily the threat posed by Assyria) as due to violation of the covenant code, which progresses through the Exile and post-Exilic period, and culminates in Jesus' rejection of all noble values. Snelson's argument is convincing and provided the framework for my interpretation of the development of the bad conscience. (Snelson, 2016)

found in the Hebrew Bible, which is itself a product of a later redaction of various written sources and oral tradition. It seems unlikely that Nietzsche, as familiar as he was with the historical-critical approach to these texts, would give much credence to the Hebrew exodus as an actual historical event, at least as described in the Hebrew Bible.

Moreover, even as described in the Hebrew Bible, the events do not reflect a slave revolt “in morality.” The slave revolt in morality is the attempt of a subjugated people to take “spiritual revenge” against their oppressors, which appears to have been motivated and orchestrated by the priesthood. (GM I:7) The exodus from Egypt describes a political revolt of sorts, where the Hebrew slaves are literally freed from captivity. Seeking literal freedom is not seeking spiritual revenge. Moreover, according to the Hebrew Bible, the priesthood did not exist until after the events of the exodus, so if Nietzsche had the Exodus in mind, it would be inconsistent with that text and unconvincing.

The mention of two-thousand year history could also be connected to the particular Jewish sect that centered around the teachings of Jesus, and developed into Christianity. This would place the end of that two-thousand year history roughly around Nietzsche’s time period. This interpretation is more intuitively plausible. The slave revolt begins with the Jews that follow Jesus and, by producing Christianity, is spread throughout the Western world over the course of two-thousand years, thoroughly influencing Western morality and culture, up to, and beyond Nietzsche’s own time.

Despite its initial plausibility, however, there are a number of difficulties for this interpretation. If the slave revolt in morality began with Jesus and his followers, then this

places the events that led to the moralization of guilt and duty and the bad conscience in the context of the Roman occupation of Palestine. I have argued that the bad conscience is produced through resentment, when the influence of “noble values” and “noble religion” leads the slaves to feel themselves under the pressure of an enormous debt to the gods, which they cannot repay. This is because, according to Nietzsche, the slaves adopted the gods of the nobles. So, if the bad conscience developed during the time of Roman occupation, this implies that it was the Jewish people that adopted the Roman gods and felt the pressure of debt to them. But this claim is completely implausible and has historical evidence against it. By this period of time, Judaism was staunchly monotheistic, while the Romans were polytheistic. Just prior to Roman occupation and under the rule of the Seleucids, the Jewish people rose up in outrage over the profanation of the Second Temple (the “abomination of desolation”), when Antiochus Epiphanes erected a statue of Zeus in the temple and sacrificed pigs on the altar. (Bright, 1972, p. 422) This does not sound like a people who would be amenable to Roman religion.

Moreover, the idea that the slave revolt began with Jesus and his followers is inconsistent with Nietzsche’s text. Nietzsche claims that Christianity, with its emphasis on God’s love, was the “triumphant crown” that grew out of Jewish “ideal-creating, value shaping, hate” and the desire for revenge. He specifically mentions Jesus as the embodied “Gospel of Love” and “Redeemer,” whose message of “blessedness and victory to the poor, the sick, the sinners” was a “seduction and detour to precisely those Jewish values and re-shapings of the ideal.” (GM I:8) Nietzsche also claims that “the slave revolt in

morality begins when resentment itself becomes creative and gives birth to values.” (GM I:10) Thus, the hate-inspired ideals and values that Jesus’ gospel of love covertly led to were created earlier. And the creators of these ideals and values are the Jewish priests, who Nietzsche describes as “the most evil enemies,” because they are “the most powerless,” whose powerlessness produces a hatred that “grows into something enormous and uncanny, into something most spiritual and most poisonous,” and seeks “spiritual revenge” against “the lords” and “power-holders,” through a radical reevaluation of their values. (GM I:7)

The Jewish priests of this period were the remnants of the Jewish nobility. Nietzsche points out that it is no exception to his rule that “the concept of superiority in politics always resolves itself into a concept of superiority of soul,” when the highest ruling caste is a priestly one. As we have seen, the priests use the terms “good and bad” to mark the distinction between “pure and impure.” Nietzsche claims that the opposition between “pure and impure” are initially marks of distinction “among the estates.” The mention of “estates” suggests a culture in which both “knightly nobles” and “priestly nobles” existed, each recognizing the other as part of the overall nobility (however much rivalry existed between them), but distinguishing themselves in terms of “knightly values and virtues” versus “priestly values and virtues.” Nietzsche also claims that one can find a later development where the sense of “good and bad” is no longer related to the estates. (GM I:6) This could suggest several possibilities. One possibility is that there are no longer any estates, but the concepts of “good and bad” continue to be used. It could also

imply a culture in which there is no longer any relationship between the estates, and this could be because only one remaining estate exists, or because the estate with political power refuses to recognize the authority of any other estate.

I want to suggest that we find both of the last two possibilities at various stages in the history of political and religious developments in ancient Israel, and that it is exactly in this history that Nietzsche locates the development of the ascetic ideal, the slave revolt in morality, and the bad conscience.

Disregarding claims about the earliest period of Israel's development, which have as their only source the accounts given in the Hebrew Bible, historians have identified a period of time in which Israel had both a priesthood and a monarchy (beginning roughly around 1000 BCE and ending 587/86 BCE). (Bright, 1972, p. 191; 330)

Here we have two different "estates," with the monarchy responsible for conquest and expansion (during the earliest years of Israel's existence), or defense against encroaching invaders (a large part of Israel's history), and with the priesthood responsible for the religious life of the community through the ritual of sacrifice conducted in the temple. With these two estates, it is plausible to claim that they would have had a distinction

between “knightly” values and “priestly” values.³⁰ Here, “pure and impure,” would serve to mark the distinction between the priests, whose main concern was with ritual purity, as a necessary condition for interceding on Israel’s behalf with Yahweh, and those whose duties as warriors tended to leave them ritually impure, in the eyes of the priests.

The Northern Kingdom of Israel falls to the Assyrians in 722 BCE. (Bright, 1972, p. 229) The Southern Kingdom of Judah is conquered by the Babylonians in 587/86 (Bright, 1972, p. 330), which puts a permanent end to any independent monarchy in Israel.³¹ The destruction of the temple puts an end to the sacrificial system and creates both a crisis and opportunity for the surviving priesthood. With the temple destroyed and the sacrificial system ended, the priests faced the threat of irrelevance. They survived by reinterpreting sacrifice as ethical and ritual purity that was achieved through prayer and study. They were able to establish their authority by reinterpreting Israel’s history as a religious history, in which the Exile was punishment for Israel’s failure to live up to the ethical and ritual purity code revealed by Yahweh to Moses. The Babylonian diaspora not only

³⁰ The Hebrew Bible can lead one to believe that the monarchy was always under the influence of the priesthood, which stood ready to dictate and judge the behavior of the monarchy in terms of “purity.” But there is good evidence (and Nietzsche was familiar with this evidence) that this priestly influence and critique was redacted into older texts and combined with oral tradition by what are known as the “priestly source” according to the “documentary hypothesis” on the origins of these texts. The priestly source is located in the time after the Babylonian Exile (Bright 68) and reflects an ascetic interpretation of Israel’s history, as Nietzsche notes in *Antichrist* 26. The tension between the monarchy and priesthood is reflected in the tension between earlier biblical sources and the priestly source. As an example, note that Samuel I: 9.1 to 10.16 & 13.3b, 4b-15 provide a favorable account of the monarchy in Israel (under Saul), while Samuel I: ch.8; 10:17-27; and 12, are highly unfavorable. Both strands are woven into the same narrative account. (Bright, 1972, pp. 187-188)

³¹ Any monarchs that ruled in Israel post-Exile were puppets of other empires.

survives the Exile, but thrives and goes on to become one of the most influential centers of Jewish scholarship.

After the Babylonians are conquered by the Persian Empire, the Jewish people are allowed to return to the Palestinian province and rebuild the temple. From this point on the only remnant of the original nobility of Israel is the priesthood, and it is during the period of exile and resettlement that followed that a sense of “good and bad” develops that no longer has any contrast *between* the estates. The terms “good and bad” are now the exclusive property of the priestly-noble conception of ethical and spiritual purity before God. It is in the post-Exilic period of resettlement that the Hebrew scriptures are redacted (Nietzsche mentions the “great literary forgery” claimed to be the discovery of a “sacred book”(A 26)) to include the priestly interpretation of Israel’s history – covenant unfaithfulness and sin – and the inclusion of legal and ritual obligations (found in Leviticus) to state the conditions of covenant faithfulness in terms of purity.³²

The history of Israel reflects key aspects of Nietzsche’s account particularly well. Nietzsche expresses admiration for the Old Testament (ignoring priestly interpretations, to be sure):

I take my hat off to the Old Testament! In it I find great human beings, a heroic landscape, and something most rare on earth, the incomparable naïveté of the strong heart; still more, I find a people. (GM III:22)

In the Jewish ‘Old Testament’, the book of divine justice, there are men, things and speeches of so grand a style that Greek and Indian literature have nothing to

³² We know that during the Exile the Deuteronomic Histories that span from Joshua to Kings II were redacted, the cultic codes of the Temple priesthood were collected and codified during, and the priestly narrative of the Pentateuch was redacted into earlier sources. (Bright, 1972, p. 350)

set beside it. One stands in reverence and trembling before these remnants of what man once was and has sorrowful thoughts about old Asia and its little jutting-out promontory Europe...the taste for the Old Testament is a touchstone in regard to 'great' and 'small' –" (BGE 52)

What impresses Nietzsche in particular is that the Old Testament presents a genuine people whose God reflected their own national and political successes and aspirations.

"Yahweh is the God of Israel and consequently the God of justice: the logic of every nation that is in power and has a good conscience about it." (A 25) Here it seems that the Jewish people are adapted happily to their community and that their aggressive instincts were internalized enough to follow the basic ethical and religious demands placed on them.

But conquest and exile in Babylon forces them to adapt to a new culture and new customs, while trying to preserve their own sense of identity. Here they face internalization of aggression again. The destruction of the temple ended cultic sacrifice. The sacrificial system functioned to provide the people with a means of internalizing their aggressive instincts. Like all religions, pre-Exilic Judaism was a system of cruelty (GM I:3), and through the practice of sacrifice Judaism provided the people with a way to direct their cruelty towards themselves (to be a genuine sacrifice, it must involve loss and suffering), and a way to both repay God and express gratitude for their destiny. (A 25)

The end of cultic sacrifice left the Jewish people without a means to internalize aggression and without a means to repay their debt to God. Thus, they also faced the need to internalize their religious instincts. They suffer from the pressure of debt consciousness and anxiety over punishment. The end of their political ambitions and their captivity is

inexplicable to them, given their belief in God's promises. They suffer from resentment, but it is not clear who is to blame for their predicament, thus resentment lacks a target and continues to mount, and threatens to erupt in anarchic violence that would destroy the Jews as a people.

Rather than accept the idea that God was no longer able to fulfill their hopes, the priests provide a solution to this predicament by altering their conception of God. God was no longer exclusively Israel's God or an expression of their national self-confidence who helps and inspires, but became a moral God who makes demands, and places moral conditions on fulfilling the covenant. (A 25) This reinterpretation includes the idea that Israel's covenant with God was unbreakable in one sense – God would forever stand ready to fulfill the covenant promise – but was conditional in another crucial sense – the covenant would only be fulfilled when the people fulfilled their covenant duties. But the duties that they must fulfill are the priestly-noble's ascetic interpretation of those duties. The people must live up to the priest's ideals of purity in order to be "good."

The priests are then able to redirect resentment by interpreting the exile as punishment for sin. The people are guilty before God because they have failed to adhere to the purity code of the priestly-nobility (good = pure; bad = impure). They only have themselves to blame for their suffering. This produces the bad conscience and guilt.³³

³³ My sense is that guilt develops on a continuum out of non-guilty debt consciousness to full-blown moralized guilt, such that it is difficult to say with certainty when the first genuine instances of guilt appear. It is easy enough to recognize guilt once Christianity arrives on the scene, but it is not clear whether the Jewish exiles felt precisely the same feeling. What they must have felt, does seem to amount to a form of proto-guilt at any rate.

The priest is then able to internalize aggression by reinterpreting sacrifice as a matter of the “heart” that takes place through prayer, study, religious devotion, and ritual and ethical purity as a means to atone for sin. In so doing, the priest encourages self-reflection, self-critique, self-contempt and self-punishment for sin. This provides a meaning for suffering and a pretext to engage in self-cruelty that satisfies the will-to-power. Like the Christian atonement this idea of atonement would provide temporary relief from the pressure of debt consciousness, since the idea that debt can be repaid through suffering and “sacrifice” is still powerful.

3.2 The Slave Revolt in Morality

One might wonder how all of this amounts to a slave revolt. The answer is that it does not, but it produces the conditions that will lead to one. Nietzsche only claims that the slave revolt *in morality* began with the Jews, and what we see in the current account is the beginning of a revolt in *values*. The priests invert the morality that existed as an expression of Israel’s sense of power, national self-identity, and self-affirmation, to produce a morality that was no longer a nation’s deepest instinct of life, but is rather, the antithesis of life. (A 25) They reinterpreted natural causes into the “moral world order” as punishment for sin. Moreover, the priest reinterprets Israel’s history in such a way that it does count as a *priestly* revolt and act of revenge against the previous moral order:

What does ‘moral world order’ mean? That there exists once and for all a will of God as to what man is to do and what he is not to do; that the value of a nation, of an individual is to be measured by how much or how little obedience is accorded to the will of God; that the *ruling power* of the will of God, expressed as punishment and reward according to the degree of obedience, is demonstrated in the destiny of a nation, of an individual ... the *priest* abuses the name of God: he

calls a state of society in which the priest determines the value of things ‘the kingdom of God’...with cold blooded cynicism he assesses nations, epochs, individuals according to whether they were conducive to the rule of the priests or whether they resisted it...in the hands of the Jewish priests the *great* epoch in the history of Israel became an epoch of decay, the Exile, the long years of misfortune, was transformed into an eternal *punishment* for the great epoch – an epoch in which the priest was as yet nothing. According to their requirements they made the mighty, *very freely* constituted figures of Israel’s history into either pathetic cringing bigots or ‘godless men’... (A 26)

So, what we encounter in the priesthood that emerges out of the Exile is the kind of spiritual revenge that is characteristic of the priests depicted in GM. The priests take revenge against the past, and those figures in the past, who denied the priesthood authority to rule.³⁴ They do this by denying Israel’s old values and promoting the priestly distinction between “pure and impure” to absolute authority.

And it seems that the priests might have been successful in their efforts for some time. The Jewish people were given a goal – the restoration of Israel – and the means to that goal – living up to priestly morality. And with the return to Judea and construction begun on the Second Temple, it might have seemed to them that they were on the way to achieving that goal.

But history proved otherwise. After the return to Judea in 539 BCE, Israel was under the rule of one empire or another from the Persians, to the Macedonians, to the Ptolemys, to the Seleucids, ending in 141 BCE. Israel had a brief period of relative independence from 141-61 BCE when the Maccabees established the Hasmonean State, but from 63

³⁴ See fn.30 on the tension within the Biblical narrative over the monarchy.

BCE until the destruction of the Second Temple and the dispersal of the Jewish people in 70 CE, Israel remained under Roman occupation and control.

Roman occupation reintroduces warrior-noble values and the conflict with priestly values resumes. The Romans introduced new customs and rules. They were extremely ruthless about suppressing groups that promoted national identity and expressed ambitions towards political independence. They looked down on Jewish culture, religion, and values with a great deal of contempt. This new context forced the internalization of the instincts once again, and with that, more suffering.

The series of successive conquests and occupation by foreign powers leading up to the Romans would have been felt by the Jewish people as a matter of God's punishment for sin. This would lead the people struggling to live up to priestly ideals of purity to see their failures as evidence that they are incapable of doing so because of their sinfulness. And it is precisely here that we see the culmination of the conflict that develops when the people (the slaves) adopt a noble code of morality (good and bad as pure and impure) and a priestly-noble conception of God (God as transcendent and holy, a moral God).

The Jewish people accept the values of "good and bad" in terms of "pure and impure," and the priestly-noble reinterpretation of God as holy and transcendent. By that standard they are incapable of following the priestly ideals of purity, which places responsibility for unfulfilled covenant promises on them. This results in a systematic frustration of the will-to-power that produces resentment. And in this instance it leads to a revolt against all noble values

Enter Jesus of Nazareth, whose radically egalitarian message rejects all noble morality. All people are equal (and equally sinful) before God. Living up to God's ideal requires a radical renunciation of all things worldly, including especially, the accepted hierarchies between "slave and free, man and woman, Jew and Greek" – and we might add, noble and ignoble. Salvation requires renunciation of the self and childlike faith in God. The Jewish priests' efforts to remain ethically and spiritually pure are as misguided as the pagan Romans' mistaken conception of virtue. This is where the slave revolt in morality, which began in post-Exilic Judaism, is brought to fruition.

The slave revolt is responsible for introducing the value oppositions "good and evil." These terms pick out the possession of certain traits and characteristics as "evil," because of their connection to the animal instincts, and others as "good," because they deny human and animal nature, and are conducive towards a morality of selflessness, self-denial, and self-sacrifice. Thus, the value terms "good and evil" are the product of an ascetic interpretation of existence. But, by the time these value terms are introduced they are applied against any type of nobility. The "evil ones" are those who seek power and dominance over others, whether these are Roman conquerors or Temple priests. The "good ones" are the weak, the poor, the humble, the sick, the oppressed, and the suffering.

So, on this account, an ascetic interpretation of reality and God is employed to promote priestly-noble values over war-like and nationalistic values sometime during the period of the exile. But accepting these values produces another religious and political

crisis for the Jewish people that leads to resentment not only against their foreign oppressors, but against what remains of their own priestly nobility. The actual revolt of the slaves is against all noble morality and gives rise to the values of “good and evil” which are opposed to any form of noble morality, whether the “good and bad” of any warrior castes like the Romans, or the “pure and impure” of priestly castes. Both warriors and priests are “evil” because they feel themselves to be superior to the slaves, and oppress the slaves through their unattainable ideals.

3.3 The Greeks and the Bad Conscience (or lack thereof...)

One important implication of my account is that Nietzsche is primarily concerned with the form of guilt that developed through the influence of Judaism and Christianity, and which continues to exist in Western culture, through the influence of ethical theories that have disguised versions of the ascetic ideal as their basis and presupposition. My thesis has been that the bad conscience, and thus the pervasive feeling of guilt, is the product of a very specific, and ascetic, religious interpretation of debt consciousness: the concepts of debt, obligation, and punishment provide the basis for this development, but they need not be interpreted in a way that leads to the bad conscience and pervasive guilt.

A further implication of my account is that pervasive guilt corrupts sovereign individuality. Through adopting the values and gods of the nobles, the slaves are burdened with the consciousness of unavoidable debt and anxiety over punishment that is pervasive guilt. But as they begin to aspire to live up to these values and ideals, and struggle to self-regulate their behavior in light of those values and ideals, the slaves

become sovereign individuals. The ability to deny oneself (or some aspect of oneself), and to affirm instead some norm, standard, or ideal, is to be capable of valuing and acting on one's values, rather than simply being motivated by momentary affect and desire. And this is exactly the case with the slaves. The ascetic ideal might be impossible to attain, but the slaves affirm the ascetic ideal and deny themselves in an attempt to live up to its standards. The slaves have become sovereign individuals.

But the slaves attained sovereign individuality through the interpretation of their suffering as guilt. When the priest identifies the animal instincts as the source of the slaves' sin and guilt, their resentment and internalized aggression are redirected towards their own animal natures. Through internalized aggression the slaves are able to engage in the activities of "burning into oneself a will, a critique, a contradiction, a contempt, a 'no'..." (GM II:18) This internal self-critique imposes a hierarchical structure on their drives as they condemn and demote the animal instincts, in favor of pursuing the ideals of selflessness, self-denial, and self-sacrifice. But as Nietzsche points out "bad conscience, the will to self-maltreatment, first supplies the presupposition for the *value* of the unegoistic. —" (GM II:18) The slaves are motivated in this project by the bad conscience and the feeling of pervasive guilt. It is a project, Nietzsche claims, in which the slave finds in God "the most extreme opposites he can find to his actual and escapable animal instincts themselves as guilt before God." Thus, the slaves' achievement of sovereign individuality is founded upon their own self-loathing and internalized cruelty and is employed to deny what is basic to human nature.

The ascetic ideal provided a meaning and purpose to human existence and human suffering: the meaning of human existence is to overcome one's sinful animal instincts and live up to the ideals of a perfect, transcendent, and holy God. Suffering is to be understood as a state of punishment (GM III:20), though "their misery is a distinction and election from God...one beats the dogs one loves the most...(GM I:14) But the ascetic ideal is an impossible ideal to attain, so its primary function is to satisfy the will to power by providing a pretext for self-inflicted suffering. The ascetic interpretation does nothing to address the source of the suffering. It only provides anesthetization of suffering through the violent affect of enjoying the pleasure of cruelty. Thus, the sovereign individuality of the slaves has been perverted into a self-reinforcing dynamic of self-cruelty, where cruelty is inflicted for its own sake.

An ascetic interpretation, however, is only one possible religious interpretation of human existence and suffering. Referring to the "holy God" of Judaism and Christianity, Nietzsche claims that:

...in itself the conception of the gods does not necessarily lead to this degradation of the imagination...there are more noble ways of making use of the fabrication of the gods than for this self-crucifixion and self-defilement of man... (GM II:23)

There are religions that do not produce the bad conscience, and so do not exploit the feeling of guilt such that sovereign individuality is perverted. These religions do, however, involve internalization of the aggressive instincts and some sense of self-cruelty.

When Nietzsche presents the development of the creditor-debtor relationship to the ancestors, he describes the development of the ancestral spirits into gods. He also claims that this development was not motivated by piety, especially in the “middle period in which the noble clans take shape.” These nobles, Nietzsche claims, “returned, with interest, to their originators, the ancestors (heroes, gods) all of the qualities that had in the meantime become apparent in them, the noble qualities.” He then tells us that “later we will take another look at the aristocratizing and ennobling of the gods,” (which, he warns us, is not to be confused with their “hallowing”). (GM II:19) Before doing so, he presents the culmination of the concept of God as the “maximum God” of Christianity. The clear implication is that the God of Christianity represents the “hallowing” of God out of piety.

Nietzsche keeps his promise to return to the ennobling of the gods when he presents the gods of the ancient Greeks. The Greeks put their concept of the gods to more noble uses. Nietzsche claims that:

For the longest time these Greeks used their gods precisely to keep “bad conscience” at arm’s length, to be able to remain cheerful about their freedom of soul: that is, the reverse of the use which Christianity made of its god. (GM II:23)

These “noble and autocratic” Greeks used their gods to deify the animal nature in human beings, such that it did not “tear itself apart, did not rage against itself!” The Greek gods were reflections of the nobles’ conception of themselves. This is what Nietzsche means when he claims that the nobles returned their own noble qualities, with interest, to their heroes and gods. (GM II:19) The Greek gods reflected human nature in all of its aspects,

whether good or bad, heroic or petty, and so were an honest and open acceptance of human nature (or at least *noble* human nature).

The stark difference between the Greek's approach and Christianity's approach is obvious. Christianity offers the "self-crucifixion and self-defilement of man" (GM II:23) over the animal instincts. It interprets animal nature as the source of sin, guilt, and suffering. It is, thus, a condemnation of the animal aspects of human nature. The Greek gods, by contrast, provide a justification for the animal aspects of human nature, by reflecting the same tendencies and behaviors of human beings on a mythical scale. The actions of the gods reveal the animal aspects of human nature to be capable of leading to bad actions, but also necessary for great actions. Thus, while the Greeks might condemn certain actions as evil or bad, it was not animal nature itself that was responsible for badness or human suffering. In the end, the gods are responsible for evil actions and human suffering.

Initially, Nietzsche suggests, the Greeks saw the source of evil in human "foolishness," "lack of understanding," and "disturbance in the head," and used these ideas to explain human "misdeeds." But, with a growing sense of their own nobility and excellence, even this explanation became problematic to them as an explanation of an 'incomprehensible atrocity and wanton act' performed by one of their peers. They wondered how this could be possible for someone of noble descent, virtue, and wisdom, who lived in the best society. The answer they gave is that "a god must have beguiled him." Thus, by placing the blame for human evil-doing on the gods, the Greeks justified

human nature – including its animal nature. The Greek gods did not take on the punishment for evil (as in Christianity), but rather, took on the guilt, which Nietzsche claims “is more noble.” This makes the gods responsible for both evil deeds and the suffering that comes from those deeds. The difference between taking on punishment and taking on guilt explains how the Greeks were able to “keep bad conscience at arm’s length.”

In the case of Christianity, it is human beings who deserve punishment for their sinful animal natures, which lead them to disobey God’s will. As we have seen, God takes on the punishment for sin in Christianity, leaving human beings with the guilt. But the connection between guilt and punishment has been deeply established in human psychology through the creditor-debtor relationship. By removing the threat of punishment from an outside source (God), the doctrine of the atonement functions to fully internalize the dynamic between guilt and punishment, leaving the task of punishment to human beings. So, the doctrine of the atonement does not actually eliminate punishment, it simply relocates and disguises it in the bad conscience.

By contrast, when the Greek committed an evil deed, it was the gods who were responsible for motivating it, and it was the gods who bore the guilt.³⁵ The evil-doer may be left to suffer punishment for his deed, but punishment was for an evil deed, not an evil person. Thus, the doer of an evil deed could understand punishment as providing repayment for the deed, and could retain a good conscience about his own nature. So, for the Greeks, it was the case that the deed “sullied the person, but punishment can provide repayment, while in Christianity it is an “already sullied human nature” that leads human beings to commit evil deeds, which requires on-going self-punishment.

Nietzsche also claims that the Greeks use of their gods allowed them to remain cheerful about their “freedom of soul.” He does not explain this claim, but it recalls his description of the sovereign individual who is “free again from the morality of custom,” who has “become free,” and has a “true consciousness of power and freedom.” (GM II:2)

³⁵ The claim that the gods “took on the guilt” may have further implications that cannot be explored at this time. The concept of guilt is rooted in the concept of debt – as something that must be repaid to the respective creditor. The problem with an evil deed in Christianity is that it is both a wrong against the community and a wrong against God, so both creditors must be repaid. This conflates ethical failure with failure to live up to an ideal of human nature. This conflation runs both ways and is the reason that religious failures – sinful thoughts, lack of devotion, ritual impurity – are felt as ethical failures that deserve punishment.

This has several implications. One has to do with the way that we understand non-moral guilt. Inasmuch as the general feeling of guilt is of pervasive anxiety and inadequacy with regard to some normative standard, it implies that non-moral guilt is a product of this conflation and may be less rational than I have argued for elsewhere.

And if the Greek gods were seen as taking on the guilt for an evil deed, it implies that, though the evil-doer might owe repayment to the community, he does not owe repayment to the gods. If this is the case, then the Greeks were able to avoid conflating ethical failures (actions that concern others) with religious failures (failures to live up to an ideal), and *vice versa*.

The further implication is that “moralization,” for Nietzsche, also involves conflating the religious and the ethical, by making God a moral being. Thus, the moralization of guilt and duty first involves interpreting God as morally perfect.

We have already seen that the sovereign individual is free from the morality of custom through possession of a conscience, and that the conscience is the result of internalized aggression and cruelty directed towards living up to the internalized norms of society. Thus, if the Greeks are sovereign individuals, they must have internalized their aggressive instincts successfully. And, if they lack the bad conscience, as Nietzsche claims, then they must have found a way to internalize aggression other than the solution provided by the ascetic priest.

Internalized aggression amounts to self-cruelty in one way or another. The fact that the Greeks remain cheerful about this requires that they remain unashamed of taking pleasure in cruelty. The problem with the slaves, in Nietzsche's account of the bad conscience, is that they have become ashamed of cruelty, despite the fact that their psychologies are driven by self-cruelty through the concepts of guilt and punishment. The resulting difference is between the Greek, who faces the use of self-cruelty honestly and openly, while the slave of bad conscience conceals it from himself. In either case, however, the suffering that is produced through self-cruelty must have a meaning and purpose. We see from the passage above that it was the Greeks' use of the gods to justify, rather than vilify, animal nature that allowed them to remain cheerful about self-cruelty and keep the bad conscience at arm's length. But, how did the Greeks' use of their gods direct aggression within and also make the suffering this produced meaningful?

The Greeks seem to have had a number of resources to provide outlets for aggression. They encouraged healthy competition between members of the community in a number

of areas: athletic contests, contests in drama and poetry, public projects to benefit the community, political engagement, and doing philosophy. But these are all ways of directing aggression towards others and were not available to the majority of the population. So it seems that the internalization of aggression must have occurred through other means.

One means of internalizing aggression is found in the effort to make oneself conform to the customs and rules of the community. The fact that the Greeks did not condemn the aggressive instincts and cruelty should not imply that they did not need to struggle against, and control the expression of those instincts. This would require aggression to be directed towards mastery over the other instincts and would involve some degree of self-cruelty. Since the Greeks did not condemn their own animal instincts, they were not engaged in a futile attempt to abolish those instincts, so their efforts at self-mastery could provide satisfaction to the will to power. Moreover, Greek religious culture was polytheistic and non-ascetic. This provided the Greeks with numerous, class-specific, role-specific, and attainable ideals to pursue as sovereign individuals engaged in the project of self-overcoming. This, too, provided a means of internalizing aggression.

Arguably the most important means the Greeks had for internalizing the instincts was through the practice of sacrifice. Nietzsche says nothing about sacrifice in his discussion of the ancient Greeks, however, so this claim will have to be defended indirectly.

Note that Nietzsche's account of the bad conscience in GM II:16 & 17 is not very specific. It does not suggest any actual historical events and presents a nearly mythical

conquest of a large group of nomadic peoples by some “blonde beasts of prey.” I take this to be Nietzsche’s general formula for the internalization of the instincts and the animal bad conscience. Nietzsche’s general formula presents and explains the effects produced when *any* originally nomadic people are conquered and subjugated into the political structure of the state. If this is the case, then it should apply to the Greeks, which implies that they had a large slave population (as we know) who underwent the internalization of their dominant aggressive instincts and suffered from the animal bad conscience.

We find support for this in the text in Nietzsche’s discussion of Hesiod and Homer. Hesiod faced the problem of trying to fit the glorious, but also gruesome and violent world of Homer into his cultural schema of gold, silver, and bronze ages. Nietzsche claims that he coped with this problem by dividing the bronze age into two ages: one that included the heroes and demigods of Troy and Thebes, to which the noble dynasties traced their ancestry, and then the very same world:

...as it appeared to the descendants of the downtrodden, plundered, mistreated, dragged-off, sold-off: an age of bronze, as stated — hard, cold, cruel, without feeling or conscience, crushing everything and covering it with blood. (GM I:11)

This indicates that the ancient Greeks went through the same process of conquest and subjugation, that resulted in noble and slave classes as we see in Nietzsche’s account of the bad conscience. Thus, the ancient Greeks had a population of slaves that went through the internalization of the instincts and suffered from the animal bad conscience.

Nietzsche’s presentation of the creditor-debtor relationship is also a general account that is meant to explain the development of economic thinking and debt consciousness.

As such, it too, should apply to the Greeks. Nietzsche connects the ancient Greeks to the noble clans of the “middle period” who were clearly conscious of their debt to the gods. (GM II:19) This means that the Greeks thought of themselves as indebted to their gods to the same degree that their community was successful. The ancient Greeks saw themselves as the pinnacle of civilization, so their debt to the gods was great. And the means to repaying the gods, and relieving debt consciousness, remains some form of sacrifice.

A sacrifice is only genuine if two conditions are met. The offering must be an actual loss to the person making the offering, so it must involve suffering. The offering must also be given up willingly. These conditions explain how the practice of sacrifice provides a means of internalizing aggression that satisfies the will to power. To sacrifice something involves willing an action that causes oneself suffering, so aggression must be internalized to overcome the aspect of the self that resists loss and suffering. Thus, to make a genuine sacrifice, the person must engage in self-cruelty and overcome internal resistance, which satisfies the will to power. In the case of the ancient Greeks, whose gods reflected their own bold and autocratic natures, and their “appalling lightheartedness and depth of desire in all destruction, in all the delights of victory and cruelty” (GM I:11), sacrifice allowed them to repay their debt to the gods and satisfy their desire for cruelty, while also celebrating their own natures without shame. Moreover, since cruelty was affirmed as a positive feature (one that provided the basis for their entire culture), even the basest Greek could find, through the cruelty required for sacrifice, at least one way of living up to the ideals represented by the gods, and affirming himself as “good.”

So, through the practice of sacrifice, the Greeks were able to offer a general means for internalizing aggression, satisfying the will to power, and preventing the development of the bad conscience. And this was possible, on my account, because the Greek gods did not represent an impossible ascetic ideal for human beings. The bad conscience developed when the Jewish people accepted the ascetic ideal from the priestly-nobles and their evaluative terms good = pure, bad = impure. As a result, they aspired to an ideal they could not attain and could only evaluate themselves as impure, hence bad. This systematically frustrated their will to power, producing resentment and further suffering, and left them vulnerable to the priestly interpretation of their suffering as punishment for guilt. The Greeks avoided this for a long period of time, because their gods did not represent unattainable ideals and offered them a way of affirming themselves as cruel, hence noble and good.

3.4 The Conscience and the Bad Conscience: “Episodic” vs. Pervasive Guilt

I argued in Chapter Two that Nietzsche identifies the cause of resentment, and thus the source of the suffering that is interpreted as guilt, in the fact that the slaves accept noble values and noble ideals (and have just shown how the history of post-Exilic Judaism fits this description). I also argued that this shows that Nietzsche is exclusively concerned with pervasive guilt, because it involves feelings of worthlessness and self-condemnation.

One remaining issue is whether Nietzsche’s concern with pervasive guilt and his call to eliminate guilt also suggests that he would eliminate what could be called “episodic

guilt,” or as some scholars prefer “locally reactive guilt” – guilt over particular instances of wrongdoing. I claim that Nietzsche’s call to the task of eliminating guilt from human psychology does not entail that he would also eliminate that feature of human psychology that connects painful feelings with wrongdoing, and so he is only opposed to “episodic guilt” inasmuch as it is connected to, and promotes, pervasive guilt. In other words, Nietzsche only objects to feeling guilt over a particular wrong, when that feeling is taken to be further evidence of one’s worthless depravity and becomes a pretext for engaging in self-cruelty.

The contrast between the ancient Greeks and post-Exilic Judaism and Christianity, provides support for this central thesis. To see why this is the case we need to consider the difference between the conscience and its distortion into the bad conscience.

I take Nietzsche’s rejection of pervasive guilt to be a rejection of the bad conscience and not the conscience itself. As we have seen, the bad conscience is an interpretation of the feeling of inescapable and escalating debt, combined with the feeling of being completely inadequate to fulfill the obligation to repay that debt. The bad conscience interprets failure to fulfill this obligation as moral failure, as culpable, and as having its source in an unchangeable feature of human nature – the animal instincts. Thus, each instance of wrongdoing is evidence of one’s flawed and sinful nature, and worthlessness in light of God’s holy standard. This provides the pretext to punish the source of the wrongdoing – oneself.

The bad conscience presumes the existence of a conscience that has become consumed with thoughts of personal and moral inadequacy. The conscience is the product of the morality of custom and clearly involves an internalized connection between failure to fulfill one's obligations and punishment. When this connection is internalized and "forgotten" (through the active faculty of suppression), it remains as a feeling pain over failing to meet an obligation (whether this is prospective or retrospective). The conscience is what allows the sovereign individual to be free of the morality of custom and relied upon to fulfill his promises and other obligations without the need for external supervision and the threat of punishment from others. And this is why I argued that the bad conscience is a distortion of sovereign individuality and the conscience. The interpretation provided by the bad conscience turns the conscience's normal and healthy response to wrongdoing into an instrument of self-torture.

It is highly implausible that Nietzsche wants to reject the conscience. As we have seen, the ancient Greeks are clearly sovereign individuals with consciences, and Nietzsche not only presents them in a completely positive light, but as *the* healthy contrast to the developments that took place through Judaism and Christianity. Thus, if Nietzsche has no opposition to the conscience *per se*, and if the conscience consists (at least in part) in an internalized connection between obligation and suffering, then Nietzsche must accept this as well. To accept the conscience as an important development in human nature is to accept the idea of "episodic guilt," as the means by which the individual motivates and evaluates his actions in light of the standards and values he

accepts. To eliminate “episodic guilt” would amount to eliminating the conscience and sovereign individuality, whereas eliminating pervasive guilt would amount to restoring the conscience to its healthy function.

I take this to explain why Nietzsche claims that the beginning of each treatise in GM is intentionally designed to mislead and to keep one in suspense until a “new truth is visible between thick clouds and amidst perfectly awful detonations.” (EH) GM II begins with a presentation of the sovereign individual’s conscience, which allows the sovereign individual to meet obligations through engaging in self-regulating behavior in light of internalized standards. And GM II ends with sovereign individuals who have a *bad conscience*, which allows them to engage in self-regulating behavior – primarily self-punishment – in light of their own internalized standard: the ascetic ideal. Thus, the distortion of sovereign individuality and the conscience amounts to redirecting the internalized psychological mechanism of self-sanctioning away from primarily motivating the individual to fulfill obligations, and towards primarily punishing the individual for being incapable of fulfilling obligations. And, while the conscience has been distorted in this way, the sovereign individual (of bad conscience) nonetheless feels a sense of power and freedom through self-punishment, and a sense of superiority over the animal nature that is punished.

Thus, to reject pervasive guilt and seek its elimination is only to reject the distortion of the conscience into the bad conscience. It is to reject the misuse of the conscience as a means of self-condemnation and self-cruelty for its own sake, in order to restore the

conscience to a healthier form and function. The healthy function of the conscience is to allow self-regulated behavior that is guided by ethical standards, and for that, the person must accept those standards as genuine and important and desire to be guided by them. If the person accepts those standards as genuine and important, then it simply follows that the thought of failing to live up them must be painful. And this is the conscience's internalized connection between wrongdoing and suffering that is "episodic guilt." So, to accept the importance of a conscience is to accept the importance of "episodic guilt."

Chapter Four: Nietzsche on Punishment and Justice

“I do not like your cold justice; and from the eye of your judges there gazes always the executioner and his cold steel.”³⁶

Introduction

One central aim of this dissertation is to apply Nietzsche’s ideas towards a comprehensive critique of punishment. This critique is based on Nietzsche’s claim that punishment is an interpretation of already existing practices and that, while punishment has been interpreted as having many different expected outcomes, the means of punishment have always been some form of intentionally inflicted suffering on the criminal or wrongdoer. The first step will be to explain what Nietzsche means by interpretation. Following that we will see Nietzsche’s presentation of the numerous ways that punishment has been interpreted to have different meanings, purposes, and expectations, and will consider what this implies about our relationship to punishment. I will then show how the concept of punishment developed and changed through the influence of the creditor-debtor relationship, explaining the implications of those developments along the way. Lastly we will turn to Nietzsche’s critique of punishment and his thoughts on justice in the community, to provide a deeper understanding of his opposition to the concept of punishment.

³⁶ Z:I:19

The puzzle that motivated this dissertation was Nietzsche's claim that, "we immoralists especially are trying with all our might to remove the concept of guilt and the concept of punishment from the world and to purge psychology, history, nature, the social institutions and sanctions of them..." (TI ###) My argument is that a solution to this puzzle is found if we recognize that Nietzsche makes a distinction between sanctions – as the community's response to crime and means to maintaining social order – and the concept of punishment, which is an interpretation of social sanctions as having the direct aim of causing suffering. Nietzsche's rejection of punishment implies that sanctions should be reinterpreted as having a different aim than causing suffering. I will defend the claim that sanctions can and should be interpreted as having the aim of containing criminal threats to the community, and that this is in fact Nietzsche's view. This implies that Nietzsche's rejection of punishment does not entail the absurd conclusion that he would leave crime unaddressed by the community.

4.1 Punishment as Interpretation

It is often the case that Nietzsche places his most important thoughts in the middle of a section or passage, and GM II is a prime example of this. In GM II:12 we find Nietzsche's presentation of the central methodological principle of his genealogical method. It is also notable that the example he uses to illustrate this principle is the practice of punishment. Nietzsche emphasizes the importance of recognizing a distinction that his rival genealogists of morality completely overlook: the distinction between the origins of something and its purpose. His claim is that:

...the cause of the genesis of a thing and its final usefulness, its actual employment and integration into a system of purposes, lie *toto caelo* [all of the sky] apart; that something extant, something that has somehow or other come into being, is again and again interpreted according to new views, monopolized in a new way, transformed and rearranged for a new use by a power superior to it; that all happening in the organic world is an overpowering, becoming-lord-over; and that, in turn, all overpowering and becoming-lord-over is a new interpreting, an arranging by means of which the previous 'meaning' and 'purpose' must of necessity become obscured or entirely extinguished. (GM II:12)

Several important points appear in this passage. The first is that any interpretation is directed at some already existing thing or practice, which the interpretation provides with a purpose, or meaning. Nietzsche's language of "overpowering" and "becoming-lord-over," suggests that he regards interpretation as motivated by the will-to-power. He makes this explicit later in the passage when he claims that "all purposes, all utilities, are only signs that a will to power has become lord over something less powerful and has stamped its own functional meaning onto it." (GM II:12) Interpreting a thing or practice is a human activity that imposes meaning or purpose on something in order to put it to use in some way.

The fact that interpretation is a product of the will to power explains why Nietzsche claims that the "development" of a thing or practice is never a matter of progress toward a goal. No interpretation is final, nor does an interpretation need to be the logical working out of a previous interpretation. A thing or practice can be interpreted and reinterpreted repeatedly to be given new meanings or purposes, and the interpretations themselves are "more or less independent processes of overpowering."

It should also be noted that Nietzsche regards the essence of interpretation as “doing violence, pressing into orderly form, abridging, omitting, padding, fabricating, falsifying.” (GM III:24) Providing a detailed explanation of Nietzsche’s view of perspectivism and objectivity is not possible here, so I will simply point out that Nietzsche does not reject interpretation, nor does his view of interpretation entail relativism with regard to truth. His perspectivism does entail that one should be alert to the possibility that any *specific* interpretation may be less than adequate to its original “text,” and may obscure or omit something important in the text, through the attempt to impose an interpretation on it. In particular, we should remain alert to the ways that the concept of punishment may have distorted features of the practices it interprets.

In the case of punishment, Nietzsche calls our attention to what is relatively permanent in it, and what is fluid. The relatively permanent aspect of punishment, according to Nietzsche, is some “practice, the act, the ‘drama,’ and a certain strict set of procedures.” (GM II:13) Here we should imagine a particular culture’s established legal and penal procedures: the drama of handing down the sentence, the procedures involved in carrying out the sentence, and the particular methods employed.

The fluid aspect of punishment is “the meaning, the purpose, the expectation tied to the execution of such procedures.” (GM II:13) Nietzsche notes that his historical methodology presupposes that:

...the procedure itself will be something older, earlier than its use for punishment, that the latter [punishment] was first placed into, interpreted into the procedure (which had long existed, but was practiced in another sense)... (GM II:13)

As far as the meaning of punishment – the end result that punishment is expected to accomplish – Nietzsche points out that in the late stages of a culture:

the concept ‘punishment’ in fact no longer represents a single meaning at all but rather an entire synthesis of ‘meanings’: the previous history of punishment in general, the history of its exploitation for the most diverse purposes, finally crystallizes into a kind of unity that is difficult to dissolve, difficult to analyze and — one must emphasize — is completely and utterly undefinable. (Today it is impossible to say for sure why we actually punish...) (GM II:13)

The confusion and lack of clarity with regard to the purpose of punishment is reason enough to call for a critique of punishment. If we began with the conviction that punishment might actually accomplish some purpose (a conviction I do not endorse), then it would still be absolutely vital that we make clear what we expect punishment to do and why we think that it can do so. (Employing punishment without any clear sense of purpose and expected result is akin to continually pushing that strange button in the rental car and hoping something good comes out of it.) This problem is exacerbated when we examine the actual purposes that punishment has been expected to fulfill.

Nietzsche offers us a list of uses that punishment has been put to historically. It is worth quoting this list in its entirety to get an idea of just how diverse the expectations attached to punishment have been:

Punishment as rendering-harmless. Punishment as prevention of further injury. Punishment as payment to the injured party for the injury, in any form (even in that of compensating affect). Punishment as isolation of a disturbance of equilibrium in order to prevent a further spreading of the disturbance. Punishment as instilling fear of those who determine and execute the punishment. Punishment as a kind of compensation for the benefits the criminal has enjoyed up to that point (for example when he is made useful as a slave in the mines). Punishment as elimination of a degenerating element (in some cases of an entire branch, as

according to Chinese law: thus as a means for preserving the purity of the race or maintaining a social type). Punishment as festival, namely as mocking and doing violence to a finally defeated enemy. Punishment as making a memory, whether for the one who suffers the punishment — so-called ‘improvement’ — or for the witnesses of the execution. Punishment as payment of an honorarium, stipulated on the part of the power that protects the evil-doer from the excesses of revenge. Punishment as compromise with the natural state of revenge, insofar as the latter is still upheld and claimed as a privilege by powerful clans. Punishment as declaration of war and war-time measure against an enemy of peace, of law, of order, of authority, whom one battles — with the means that war furnishes — as dangerous to the community, as in breach of contract with respect to its presuppositions, as a rebel, traitor, and breaker of the peace.— (GM II:13)

Nietzsche points out that this list is incomplete and that punishment is “overladen with utilities of all kinds.” (GM II:14) Even so, the length of the list and the diversity of expectations is telling. Human beings have expected punishment to accomplish so many different results that, if were not for the fact that these different purposes have obscured and replaced previous purposes as they came into and out of existence over long periods of time, it would strain credulity to think that anyone could believe that one practice could successfully accomplish all of these purposes. But as Nietzsche points out, human thinking has difficulty moving (and I presume this implies changing as well), but once it begins it continues relentlessly in the same direction. (GM II:8) The concept of punishment is direct evidence of this phenomenon, and perhaps reveals that we are more attached to the practice itself, than to any result that we expect it to produce.

According to Nietzsche’s account, punishment is an interpretation of already existing practices, which is a matter of some will to power imposing a meaning and expected result on those practices.

The next step is to understand how the concept of punishment developed into its original non-moral form, and how that original form retains some of its shape in the ensuing historical reinterpretations of the purpose of punishment. In the next section I turn to the practices of early human communities and how they were interpreted as punishment through the creditor-debtor relationship and economic thinking. This examination will allow us to identify what remains essential to punishment in any of the ways in which it has been interpreted. The most fundamental practice that is interpreted is the community's response to members who break the rules of that community, which in turn, is a reflection of the way that early communities treated their enemies. Thus, it is the community's response to a perceived threat or source of harm that is the underlying fact that is interpreted as punishment. In the end, I will argue that this underlying fact is a phenomenon that any community must address in order to ensure social stability, but that the response to "crime" can be interpreted as a sanction that need not be thought of in terms of punishment.

4.2 Punishment in the Creditor-Debtor Relationship

We have seen that the creditor-debtor relationship develops as early human communities grow larger and become more sophisticated and require new ways of managing exchanges of goods and services. The creditor-debtor relationship itself is a further development and interpretation of earlier and more primitive forms of exchange, trade, and barter between members of a community.

In the earlier and more primitive exchanges, it is plausible to expect that those who felt “cheated” or “let down” by the other party to the transaction would retaliate in anger over this outcome. This would certainly be the case where the other party’s failure to reciprocate endangered the other member or the community as a whole. In these cases, the response of the community is to banish the member who threatens the community.

Economic transactions become more complicated when a community’s size and stability allows for increased division of labor and specialization, which results in the accumulation of surplus goods. With this development some members are in a position to provide loans to other members based on a promise to repay the loan. And with this set of circumstances we have the creditor-debtor relationship.

It seems likely that the earliest instances of these creditor-debtor transactions were less formal and less thought out than their later forms. Individuals made agreements which were either kept or broken, for one reason or another. Broken agreements produced anger over losses and led to retaliation. This creates a crisis in the community. The creditor’s retaliation against another member is a threat to community stability, and perhaps also to the authority of the community’s leaders. But the debtor’s failure to repay the loan threatens the stability of the economic transactions that have become a necessary feature of the community’s life. The creditor’s wealth is recognized as a source of benefit to the community, so the creditor must be provided with assurance that loans will be taken seriously and repaid. The community needs to provide those assurances in a way

that will preserve the stability of the community and the economic practices that have become central to it.

Transactions between creditors and debtors become a matter of a contract that is enforced by the community. The contract includes provisions for appeasing the creditor's anger in the case of material losses, by permitting retaliation against the insolvent debtor. Retaliation provides the angry creditor with the pleasure of making the debtor suffer, and also provides a feeling of social elevation, since it allows the creditor to engage in an activity that is reserved for the rulers of the community, and provides him with an opportunity to hold another person in contempt. The feeling of satisfaction derived from causing suffering and holding in contempt outweighs the dissatisfaction felt over the material loss, and thereby compensates the creditor for that loss.

Authorizing retaliation against insolvent debtors is a crucial step in the development of punishment, but it is not yet complete. The more significant development is the equivalence principle. Through the economic thinking required to successfully engage in transactions of goods (gauging values, determining equivalents, setting prices), the parties to these transactions eventually arrive at the idea that equivalences can be found for every material good, "every thing has its price." This thought is then extended to include the physical suffering that the debtor endures at the hands of the creditor. If material goods can satisfy a debt, and causing a debtor to suffer can satisfy a debt, then, so the thought goes, there must be an equivalence between the loss of material goods and the degree of suffering the debtor must endure. With this development we have a non-

moral concept of punishment that explicitly connects debt with an expected and authorized amount of suffering. Over time, this establishes a deep psychological connection between indebtedness and suffering.

We should pause here to note that the development of this non-moral practice of punishment is already strange and questionable. The idea that suffering can provide compensation for a material debt is in fact bizarre. The suffering of the debtor does nothing to restore the creditor's material losses. Punishment as compensation relies on the fact that aggressive human beings naturally retaliate in anger over losses or injuries, and take pleasure in inflicting suffering on those who cause those losses or injuries. The creditor-debtor relationship relies on the questionable equivalence between debt and suffering to justify instinctive and angry retaliation as an appropriate response to an injury or loss. Thus, the practice of punishment begins as an accommodation to human anger and cruelty.

The creditor-debtor relationship produces further developments to the equivalence principle. The fairly natural idea that equivalence can be found between material goods, led to the idea that there are also equivalences between material debts and physical suffering. This idea is extended beyond material and economic debts to include other actions that anger the community and its members: breaking the rules and customs of the community. Thus, the material equivalence principle is extended to include other forms of harm and injury and becomes "every thing has its price; everything can be paid off."

With this development, as we have seen, the concept and practice of punishment is interpreted onto the community's response to those who break the rules – the “criminals” of the community. Individual members are interpreted as having made a contract with the community, and in exchange for the benefits advanced by the community, each member owes obedience to rules and customs. Breaking the rules is interpreted as a breach of contract, for which the community is entitled to compensation. I argued earlier that this interpretation is first applied to banishment as a form of punishment that justifies the community's response as “fair,” because it is means of depriving the insolvent debtor of the goods he illicitly gained from the community. Later, as the equivalence principle gains more traction, banishment as a form of punishment is seen as too extreme for many violations, and is replaced by other methods of making the criminal suffer that are deemed equivalent to the specific crime.

We should pause again to take note of the way that the concept of punishment has been altered in this new context. While the equivalence principle did serve to moderate the community's treatment of the criminal, by replacing banishment with less severe responses, it also established several questionable connections.

The creditor-debtor relationship interprets obedience to the rules as the means by which one repays one's debts to the community. This means that crimes, by their very nature, and independent of any actual harm done, are interpreted as failure to repay one's debt. And according to the logic of the creditor-debtor relationship any failure to repay a debt requires suffering through punishment.

The economic thinking of the creditor-debtor relationship seems ill-suited to address crime in the abstract. When applied to crime the concept of debt starts to become ambiguous. In the relationship between individual buyers and sellers, there is a specific material debt that is owed and failure to repay that debt can be calculated as a specific loss or harm to the creditor. Some crimes result in material losses or physical harms, and in these cases the idea of adequate compensation can be brought to bear. Property loss or damage can be calculated in economic terms; medical bills can be calculated; loss of function and opportunity through physical injury, whether temporary or permanent, can also be assessed economically. Here we can arrive at some sense of what the criminal might owe the victim by way of repayment.

But some crimes do not result in material losses or physical harms. Criminals are caught in the act, prevented from following through on the crime, or made to return what they have stolen, etc. Yet their actions are still regarded as criminal and a failure to repay a debt. The idea that a crime is, in and of itself, an unpaid debt is not easily translatable into economic terms. With no material losses or physical injuries to calculate there is no way for the criminal to provide material compensation. But the community still feels anger towards the criminal. Thus, it is nearly inevitable that suffering to pay off a crime appears as the only solution the creditor-debtor relationship can provide.

This indicates that the response to crime is a product of the outrage and fear of the community. The notion of an equivalence in suffering relies on the intensity of the injured party's anger and feeling of injury, or in the community's judgment concerning exactly

how angry or injured the average victim of such a crime is entitled to feel. Punishment as an interpretation of the community's justified response to crime then serves as a pretext for those who are angered over crime to inflict suffering on the criminal, even in cases where there is no actual injury or loss.

We see similarly strained economic thinking in other aspects of the response to crime. In the relationship between individual buyers and sellers, the debt retains its original material basis. Collateral can be handed over; loans from other sources could be used to provide repayment; or a friend could repay the debt. Material compensation can, in principle, always be provided. But, in the relationship to the community, where following the rules is the form of repayment, this option is lost once a rule has been broken. It is not possible to go back and follow the rule that was broken. Following the rules in the future is paying precisely what is owed then, and so cannot provide repayment for a rule that was broken in the past. So, again, it seems that repayment of the debt must inevitably be provided through suffering.

The fact that suffering has been made central to the community's response to crime has several ironic results. As we will see, Nietzsche claims that the practice of punishment held back the development of feelings of guilt for the longest time. If the community has an interest in its criminals coming to recognize certain actions as not to be done, then it seems that punishment has worked against that goal. In the same vein, since Nietzsche claims that punishment only serves to make criminals colder, harder,

more calculating, and prudent, it seems that its use has largely served to produce better criminals, rather than better members of society.

The most important point to clarify at this stage is the essential feature of punishment. As we have seen, Nietzsche points out that punishment has been interpreted as having many different aims throughout history. At this point in time it is nearly impossible to say what we expect punishment to accomplish. But each interpretation of punishment employs the same means to whatever it is it seeks: the suffering of the criminal. Through the influence of the creditor-debtor relationship and the interpretation provided by the concept of punishment, the idea that has been firmly established in human psychology is that inflicting suffering intentionally on a criminal is justified, the criminal deserves to suffer, and indeed must suffer for his crime. But this is only an interpretation that masks the fact that the underlying motivation is to retaliate and enjoy the pleasure of making suffer.

4.3 Punishment Does Not Awaken Moral Guilt

After listing the various purposes that have historically been attached to punishment, Nietzsche argues that one widely assumed purpose can be definitively eliminated. This is the idea that the essential purpose and usefulness of punishment is to awaken “in the guilty one the feeling of guilt.” (GM II:14) Nietzsche claims that even in his time, let alone for the long prehistorical period of human development, this idea is psychologically implausible. Beginning with the prehistorical period of human development we will see why Nietzsche thinks this is the case.

Nietzsche claims that, during the long prehistory of human development, it was through punishment that the “development of the feeling of guilt has been most forcefully held back — at least with respect to the victims on whom the punishing force vented itself.” (GM II:14) His reasoning is that the criminal witnesses the judicial system engaging in the very same activities for which they condemn the criminal, but because they are committed “in the service of justice,” they are approved and committed by the authorities with a good conscience.

Robbing, overpowering, slandering, taking captive, torturing, murdering as displayed in the various kinds of punishment — all of these thus actions his judges in no way reject and condemn *in themselves*, but rather only in a certain respect and practical application. (GM II:14)

Nietzsche appears to suggest that the practice of punishment held back the feeling of guilt for several connected reasons. One reason is the fact that the actions of the punishers are done with a good conscience. But the actions of the criminal and the actions of the “authorities” have the same effect – they cause suffering – so the problem is not the actions themselves, or the suffering caused, but the person who commits those actions.

We have seen that in the earliest communities the “authority” to do violence within the community was reserved for the alpha figures who enforced obedience. When the creditor-debtor relationship developed, a new source of power and social standing was introduced: material wealth. As this source of power was recognized as important to the community, the privilege of punishing was extended to include creditors who have not been repaid. Thus, the right to punish others is a matter of those who have power within the community. This fact prevents the criminal from seeing his own actions as wrong in

themselves, but rather, as forbidden to some on the basis of their lack of power and lack of social standing.³⁷

Nietzsche also claims that the judges and punishers perform these actions “based on principle, without even the excuse of an emotion.” The criminal, this seems to suggest, can understand doing violence out of anger, stealing out of desire, or any number of other actions prompted by an instinctive emotional response. But witnessing his own punishment carried out on the basis of a principle, and for no perceivable material gain, can only prompt the thought that his punishment is motivated by the desire to cause suffering. And again, this would not reinforce the thought that causing suffering is wrong, but only the thought that it is a right for some, but is denied to others.

Moreover, Nietzsche claims that the actions of the judges and punishers themselves expressed nothing “that suggested one was dealing with a ‘guilty one’.” But rather with an instigator of injury, with an irresponsible piece of fate.” (GM II:14)

Here, I take Nietzsche to be reminding us that the practice of punishment was originally, and for the longest period of human development, not a moral practice. It functioned to provide creditors with repayment for defaulted loans, and may have also functioned as a deterrent, but it did not include the thought that the criminal was culpable and deserved punishment for that reason. The criminal seems to share this view as well. The criminal’s response to punishment includes no more “inner pain” than one might feel

³⁷ This explains why those who stand lower in the social order value the opportunity to punish so highly, feel elevated by it, and feel that they are participating in a “right of lords.” (GM II:5)

over “the sudden occurrence of something unanticipated, of a frightful natural event, of a plummeting, crushing boulder against which one can no longer fight.” (GM II:14)

Punishment cannot awaken the feeling of guilt in the criminal, because the criminal is precisely the person who lacks a conscience. Moreover, punishment cannot create a conscience for the individual, such that the feeling of guilt would follow. We can see why punishment fails in both cases by examining the contrast between the criminal and the sovereign individual, and what it means to have a conscience.

Nietzsche’s example of a person with a conscience is the sovereign individual. The sovereign individual’s conscience consists in the internalized norms of his society, internalized aggression, and the internalized connection between disobedience and punishment. Internalized aggression becomes a mechanism of self-sanctioning in the sovereign individual, and is used as a means to following the rules and norms of his society. Internalized aggression is the cause of the “pang of conscience” at the thought of breaking the rules of the community. But the conscience is the product of internalized aggression being brought to bear against the expression of any instincts that might lead to breaking those rules. Thus, following the rules of society, by denying expression to other instincts, satisfies the sovereign individual’s aggression and will to power. The important point here is that this dynamic is internal to the sovereign individual. When the sovereign individual denies expression to, or commands obedience from the other instincts, he experiences a feeling of power. By the same token, if the sovereign individual were to break the rules of the community, his conscience would “punish” him for doing so, and

he would experience a feeling of power through punishing and judging himself to have failed to live up to the standards of ethical behavior that he accepts.

Unlike the sovereign individual, the criminal does not satisfy aggression through obedience to the norms of society, but rather, through the direct expression of aggression. (Perhaps this is because the criminal is a more powerful, active type of person, not unlike the warrior-nobles, and has not yet been punished into submission.³⁸) The criminal lacks a conscience, unlike the sovereign individual, so has not internalized the norms of society, and has not internalized aggression towards the aim of following those norms. Thus, there is no internalized struggle between the criminal's aggressive drives and the other drives in an effort to become obedient to the rules. With no internalized struggle, there is no internal suffering, and the criminal does not need the ascetic priest's interpretation of that suffering as guilt.

When the criminal is punished for his crime, suffering is imposed on him by others. If the punishers are sovereign individuals, they may intend for the criminal to learn the importance of the rules, by showing him the connection between failing to obey the rules and suffering as a consequence, but the criminal does not see the connection in the same way. When punishment is externally imposed, the criminal does not feel a connection between following the rules and the satisfaction of his will to power through overcoming his own drives and instincts, as the sovereign individual does. The criminal only sees the

³⁸ Nietzsche claims that "The criminal type, this is a strong type of person under unfavorable conditions, a strong person made ill. He needs a wilderness, a nature and form of existence that is somehow freer and more dangerous; this is where all the arms and armor of a strong person's instincts *rightfully belong*. (TI, Skirmishes 45)

punishers satisfying their own will to power through inflicting suffering on him. This simply reinforces his own disposition to think that the will to power is satisfied through the direct expression of aggression. Thus, the practice of punishment does not encourage the internalization of aggression as an internalized means to following the rules, it does the opposite. It teaches the criminal the importance of becoming more cautious and cleverer when it comes to avoiding punishment, but it does not teach him the importance of avoiding crime *per se*.

Moreover, the criminal can identify the cause of his suffering immediately in those who punish. For Nietzsche, it seems that it is natural for human beings not to seek the cause of their suffering in themselves, so this situation provides little incentive for the criminal to think that the source of his problem lies within himself. Instead, the criminal is provided with a source of suffering that can be opposed and resisted, precisely by proving himself strong enough to endure their punishment. So, provided punishment does not break the criminal completely, it does not encourage self-reflection or awaken guilt, but rather, increases the criminal's power of resistance. This means that punishment actually increases the criminal's will to power through learning to endure suffering.

The criminal might learn something from punishment, but none of what he learns contributes to his moral development or the awakening of guilt. The criminal may learn that it is a better policy to avoid acting in certain ways since they have a higher likelihood of leading to suffering, but, as Nietzsche suggests, this may only serve to make the criminal more cautious, more prudent, mistrustful and secretive, and a better criminal.

Thus, punishment might make the criminal fearful of acting in certain ways, but it does not make the criminal feel guilty over those kinds of actions.

The other thing the criminal might learn is that punishment provides payment for the crime. The debt model of punishment provides the criminal with a meaning for his suffering that runs counter to any aim of awakening guilt within. Suffering is the price one pays for crime (if one is caught). One merely needs to endure suffering for the debt to be repaid – one need not take into account harms done to others, nor have any thought of apologies or reparations owed. Thus, punishment as repayment for a crime only serves to direct the criminal's attention towards his own situation and does nothing to suggest to the criminal that the problem with his actions are that they cause harm to others.

4.4 Punishment and Justice

To fully understand and appreciate Nietzsche's opposition to punishment we will need to examine his views on justice and its relationship to resentment. Resentment is behind the slave revolt in morality, it characterizes the motivation of the priestly class, is central to the ascetic ideal, and is a central cause of the development of the bad conscience and guilt. The role of resentment in the development of morality has been widely covered by Nietzsche scholars, so the claim that Nietzsche is opposed to resentment itself should be unsurprising. But Nietzsche's views on justice have received less attention from the scholarly community and his claim that justice is opposed to resentment is important and deserves more consideration.

Nietzsche's thoughts on justice are presented as an explicit rejection of the attempt by some scholars (Dürring in particular) to locate the origins of justice in the reactive affects and resentment and to, as he puts it, "hallow revenge under the name of *justice*." Thus, what Nietzsche is rejecting is a "revenge model of justice." To ground justice in resentment and revenge is obviously an attempt to provide a justification for punishment as revenge, so we can also claim that Nietzsche is opposed to the revenge model of punishment.

Nietzsche's objection to the revenge model is that it treats justice as nothing more than a development of the feeling of injury. Nietzsche counters this claim by presenting its "stark reversal" in his claim that "the last ground conquered by the spirit of justice is the ground of reactive feeling!" (GM II:11) This places justice in direct opposition to the goals of reactive feelings like resentment. We can begin to understand Nietzsche's position by examining the connection between the feeling of injury and resentment. But to do this, we need to recall an important distinction concerning resentment, and consider the context in which resentment originates.

Resentment is a reactive feeling that occurs when there is some frustration to the person's will to power, i.e. resistance to that person's efforts to effectively impose their will on the world in some way. This frustration is a form of suffering, which produces anger over that suffering and the desire to retaliate against the cause of that suffering. But, as we have already seen, resentment can occur in episodic form, or it can become a stable and defining feature of a person's psychological make-up.

Ressentiment occurs in episodic form when the frustration of the person's will to power can be overcome by discharging resentment externally. This can be accomplished through direct action that addresses the source of frustration. This is what Nietzsche has in mind when he describes the warrior-nobles as going straight ahead toward danger and the enemy, or taking revenge against another noble. (GM II:10) But discharging resentment can also be accomplished by directing it into other activities that are not the actual cause of resentment. This is what Nietzsche has in mind when he describes the warrior-nobles as recovering from the tension of living within the constraints of the community by directing their aggression towards the world outside of the community. (GM II:11)

Ressentiment becomes a stable and defining feature of a person's psychology when it has no opportunity to be expressed and discharged externally. In this case, the frustration of the will to power is systematic, and with no means to discharge frustration and anger against the cause of suffering, resentment festers and "poisons" the individual. (GM II:10) When this occurs we have the "human being of resentment." We see this in the case of the priests of GM when they experience frustration of their desire for political power in their rivalry with the warrior-nobles. (GM II:7) We also see it the slaves who, on my account, experience resentment when they perceive themselves as utterly inadequate to repay their debt to the gods and avoid anxiety over punishment. In these cases, resentment is characterized by a sense of powerlessness, anger, hatred, and desire for revenge in some form (even if the revenge is only imaginary, or in effigy).

It is clear that Nietzsche views the “revenge model of justice” as a product of the human being of resentment and not episodic resentment. For one reason, as soon as he mentions resentment in connection with the “revenge model” he directs the reader’s attention to the anarchists and anti-Semites of his culture, among whom “this plant now blooms most beautifully.” Even more strongly, he claims that the effort to scientifically appraise the reactive affects (such as hate, envy, ill will, suspicion, rancor, revenge) is motivated through the spirit of resentment itself. (GM II:11) This reference to the “spirit” of resentment indicates that these are “human beings of resentment.”

The human being of resentment can be characterized by the reactive affects that Nietzsche mentions (hate, envy, rancor, revenge), but the true source of these affects is found in the feeling of powerlessness. The feeling of powerlessness over one’s inability to impose one’s will directly on the world produces these affects. More importantly, the feeling of powerlessness and the affects it produces color the way in which the human being of resentment interprets the world. And it is here that we can understand Nietzsche’s claim that the “revenge model” treats justice as the further development of the feeling of injury.

According to Nietzsche, the necessary direction of resentment is toward the outside world and is a “reversal of the value-establishing glance.” This has several implications for understanding the role of justice in opposition to resentment. The first is consistent with the claim that Nietzsche thinks of human beings as not naturally inclined to locate the cause of their suffering within themselves. The presence of resentment exacerbates

this tendency. The human being of resentment only looks to the outside to locate the cause of suffering, and does not consider that the suffering felt is not exclusively the fault of some other person. To put it more clearly, the human being of resentment already suffers from internalized aggression, and:

Every sufferer instinctively seeks a cause for his suffering; still more precisely, a perpetrator, still more specifically, a guilty perpetrator who is receptive to suffering — in short, some living thing on which, in response to some pretext or other, he can discharge his affect...” (GM III:15)

An injury may cause some additional suffering, but the human being of resentment will treat the injury as a pretext to discharge resentment, and will blame the agent of that injury for the totality of the suffering that they feel. Resentment over-exaggerates the severity of the injury in order to grant itself the possibility of full discharge. This is why Nietzsche describes resentment as “that most dangerous blasting and explosive material,” that threatens to destroy the herd. (GM III:15) When resentment has no target it seeks any pretext to make someone a target. The problem that justice must address is when resentment finds a plausible pretext, such as a crime.

The second implication follows from resentment’s nature as a “reversal of the value-establishing glance,” and is resentment’s tendency to misconstrue and falsify its target into a monstrous caricature. Values were first created, according to Nietzsche, out of the spontaneous and “triumphant yes-saying to oneself” of the nobles, who only sought their opposite in order to engage in self-affirmation “still more gratefully and more jubilantly.” The primary value in noble morality is “good,” which the nobles use to affirm themselves and their happiness, “bad” is an afterthought that, at worst, misconstrues its targets

through carelessness and lack of familiarity. Nietzsche claims that in the noble manner of evaluation there is “even too much of a feeling of cheer in oneself, for it to be capable of transforming its object into a real caricature and monster.” (GM II:10)

The values created through resentment – the slave values of “good and evil” – were only possible as a reaction to the original noble values (good and bad). The creation of values through resentment was “from the ground up, reaction” and emerged only as a “no” to what was “outside,” “different,” and “not-self.” (GM II:10) Resentment by its very nature is a reflexive rejection of something already existing – it requires something already existing – and is lacking in subtlety, complexity, or nuance. When resentment rejects something, it tends to reject everything about it without discrimination.

We see this lack of subtlety reflected in the slave revolt’s reaction to the nobles, which was not merely a rejection of their values, but was also a rejection of the nobles’ attributes and happiness. The slaves (who are human beings of resentment) treat “health, being well-formed, strength, pride, a feeling of power” as things for which the nobles will someday have to atone. (GM III:14) Nietzsche also describes the slaves as a “whole trembling earth of subterranean revenge, inexhaustible, insatiable in outbursts against the happy, and likewise in masquerades of revenge, pretexts for revenge.” He goes on to say that the final triumph of the slaves would be to “succeed in shoving their own misery, all misery generally into the conscience of the happy: so that the happy would one day begin to be ashamed of their happiness.” (GM III:14)

Thus, the human being of resentment tends to feel an injury received as the cause of the totality of his suffering. Combined with the desire for revenge this leads to the distortion and caricaturing of the agent of that injury. Resentment appraises its targets falsely and with prejudice, and turns the “enemy” (the agent of injury) into “the evil enemy,” “the evil one.” Moreover, in caricaturing the enemy as “evil,” resentment does not simply judge an action to be wrong, or object to the attitude or motivation of the agent of injury, but goes on to reject the enemy entirely, including features of the enemy that are morally irrelevant. (I take this to be why Nietzsche describes the slaves as “deceitful” and “malicious.” (GM III:15))

All of this is in the service of providing a pretext for revenge, in some form, that will allow resentment to discharge itself entirely. And as we have already seen, the effect that is sought here (unconsciously) is to anesthetize the person’s suffering through violent affect. Thus, resentment seeks to make something living suffer in order for the person of resentment to feel relief from suffering.

This is precisely what justice opposes. Being just, for Nietzsche, is a positive way of behaving. Acting justly, he claims, is a “high clear objectivity — that sees as deeply as it does generously.” This is contrasted with being “merely cold, moderate, distant, indifferent.” (GM II:11) This seems to connect the “spirit of justice” with a kind of intellectual honesty (though given Nietzsche’s comments, and his perspectivist approach to knowledge, this would be more than a merely intellectual affair, and would include insights gained from considering different affective points of view). To say that justice is

not “cold, moderate, distant, indifferent” indicates that it is engaged with, rather than detached from the objects it appraises – that there is concern involved. To describe justice as a high, clear objectivity indicates that, while it might take into account the affects at play, it does not view the matter from one perspective alone, but rather, views the matter from the perspective of a larger concern. This perspective is the functioning of the community itself.

Nietzsche claims that “justice” and “injustice” only come into existence when a community is strong enough to establish the law, and makes an imperative declaration of “what in general is to count in its eyes as permitted, as just, what as forbidden, as unjust...” The establishment of the law aims at training the community (even the injured party) to appraise crimes in an ever more impersonal fashion. In this, the law seeks to push back against the predominance of affective reactions and attempts to divert the feeling of its members away from the immediate injuries caused by crime as “acts against the law, as rebellion against the highest power itself.” (GM II:11)

Nietzsche’s contrast is with Dühring’s view, where injustice begins with the act of injuring, and “sees only the viewpoint of the injured one.” Nietzsche claims that this is “devoid of sense.” He claims that, life itself acts in an essentially injuring, violent, pillaging, and destructive manner, so injuring in itself cannot be unjust, and takes this to show that the conditions of justice are always “exceptional conditions” that are “partial restrictions of the true will of life – which is out after power...”

Nietzsche's idea seems to be that individual human beings naturally seek power for themselves – they attempt to impose their own wills on the world and shape it to suit their own needs and purposes. The will to life (synonymous with the will to power, it seems) essentially involves acting in an injuring, violating, etc. manner, but in a community, these actions must be subordinated to the end of creating a greater unit of power: the community itself.

Thus, it seems that while Nietzsche rejects the social contract theory of society, there are aspects of it that exist in his own account. Working together as a community provides individuals with more power, and through that greater advantages and benefits. But to be a community, and have the power that provides greater advantages, requires the individual members to forego certain pursuits of power, since they undermine the stability of the community. Thus, there is a kind of contractualist arrangement with regard to the laws of society. The laws reflect the conditions that must obtain in order for the community to function as a community, and articulates the requirements that must be met by the individual in order to participate in society and receive the advantages it offers.

I take Nietzsche's thought that the individual must subordinate their own pursuit of power toward the end of the community as a greater unit of power to indicate two threats to the community, both of which occur when the individual refuses to subordinate their own pursuit of power towards the end of the community. The first is the pursuit of power that leads to crime, the other is the pursuit of power through revenge against the criminal.

Crime obviously affects the community adversely, but from the point of view of justice, it is less a matter of the immediate injuries that are caused to the victims of crime, and more a matter of the destabilizing effects that crime has on the community. A particular crime may cause injury to an individual, but left unaddressed the crime creates fear and distrust within the community and undermines the conditions for cooperative endeavors. The community can only exist through cooperative endeavors, so unaddressed crime threatens the basis for its existence.

But unaddressed crime does more than create fear and distrust, it also creates resentment and the desire for revenge. Resentment driven revenge is just as much a threat to the community as unaddressed crime. As we have seen, resentment tends to falsify and exaggerate the extent of the injury received and the nature of the criminal, and seeks a pretext to discharge itself fully on the criminal. This leads to excesses of violence that easily outstrip the severity of the crime. Excessive responses to crime tend to be met with excessive counter-responses that all too readily lead to the blood feud, the vendetta, and general violence (especially if we take the community, as Nietzsche certainly does, to be largely populated by individuals who suffer from the internalization of the instincts). Each act of violence creates a pretext for violence that threatens to consume the entire community and destroy it. Thus, those that administer justice in the community must address crime, but must also prevent crime from being addressed by the injured party, or the general outrage of the community as a whole.

This is why Nietzsche claims that the administration of justice has always been found in the sphere of the active, strong, and aggressive human beings, and why he counters Dühring's view of "justice as revenge" with the claim that justice is:

...precisely the battle against reactive feelings, the war against them on the part of active and aggressive powers that have used their strength in part to call a halt to and impose measure on the excess of reactive pathos and to force a settlement. (GM II:11)

The active, strong, aggressive humans are, for Nietzsche, always the source of values and creative activity. It was this type of human being that created the community in the first place, and it is this type of human being that has the aim of preserving the community as a whole. The weak and vengeful may speak in terms of "justice," but what they really seek is revenge and anesthetization of their own suffering. They in fact subordinate the good of the community to their own ends of seeking power over others. But the strong and active have a concern for the community itself and have subordinated their aggression towards maintaining the community – they have the true need for justice. Moreover, these active and strong individuals have no need to appraise the criminal or the crime falsely, and so are in a position to provide justice for the community. Thus:

Everywhere justice is practiced and upheld one sees a stronger power seeking means to put an end to the senseless raging of resentment among weaker parties subordinated to it (whether groups or individuals), in part by pulling the object of resentment out of the hands of revenge, in part by setting in the place of revenge the battle against the enemies of peace and order, in part by inventing, suggesting, in some cases imposing compensations, in part by raising certain equivalents for injuries to the status of a norm to which resentment is henceforth once and for all restricted. (GM II:11)

Understanding the way in which justice is opposed to resentment helps to further clarify several of Nietzsche's comments regarding the growth of the community and changes to its penal codes.

Nietzsche claims that as a community grows and has an increased sense of power and confidence, its penal law always becomes milder and "every weakening and deeper endangering of the former [the community] brings the latter's [penal laws] harsher forms to light again." (GM II:10) The strength of the community is one factor in this development. When the community attains a certain level of power, it is simply a matter of fact that certain crimes no longer threaten its stability and continued existence. Initially, it may be a case where this is the perspective of those who administer justice, who then begin to impose restrictions on responses to crime. But Nietzsche also claims that one aim of establishing the law is to teach the community's members to see crime from an ever more impersonal perspective. So, it seems that as the community grows stronger and the rule of law is more firmly established, an impersonal appraisal of crime becomes more of a norm for the community, and less resentment is felt over individual instances of crime.

Nietzsche also claims that "it would not be impossible to imagine a consciousness of power in society such that society might allow itself the noblest luxury there is for it — to leave the one who injures it unpunished." (GM II:10) This claim can appear puzzling, if not disturbing. It seems to suggest a society that leaves crime completely unaddressed, since Nietzsche goes on to have that society say, "What concern are my parasites to

me?...Let them live and prosper: I am strong enough for that!” He follows this with the claim that “the justice that began with ‘everything can be paid off, everything must be paid off’ ends by looking the other way and letting the one unable to pay go free.” Nietzsche claims that this is the “self-cancellation of justice” that goes by the name of “mercy, which “remains the privilege of the most powerful.”

The idea of a society that ceases to address crime is both objectionable and implausible. It is objectionable if Nietzsche is imagining a society in which the strong, wealthy, and protected rulers have concluded that crime is no longer a threat to the community’s survival as whole, and so take a magnanimous approach to crime by ignoring the injuries or damages that the community’s individual members incur through crime. In this case, the “privilege of the most powerful,” which allows them to “look the other way” amounts to complete disregard for the victims of crime, in favor of mercy towards the criminal, perhaps solely because the most powerful feel power through letting the criminal go free.

Moreover, if this is the scenario Nietzsche has in mind, it is completely implausible to think that a society could survive while leaving crime unaddressed. Sooner or later, this approach would lead to the fear, distrust, and resentment that justice is meant to prevent. This would undermine the conditions that are necessary for a society to exist and function, and it would soon crumble, or revert to harsher treatment of its criminals, as it began to gain a sense of its own dissolution.

This interpretation, however, ignores several different facets of Nietzsche's claims. Note first, that Nietzsche claims that it is not impossible to imagine *a society* that is so conscious of its power that it leaves crime unpunished. This suggests that it is not merely the rulers who have this sense of power, but society as a whole that feels powerful enough to leave crime unpunished. This should indicate that the entire culture of the society has rejected punishment, not just those in political power.

Nietzsche's description of justice is of a "high clear objectivity — that sees as deeply as it does generously," and he describes the mercy shown to the criminal as the "self-cancellation of justice." I take the "ideal community" he imagines as "not impossible," to be one that has become objective enough to appraise the problem of crime and justice clearly, and has come to see punishment itself as a problem.

As we have seen, in the "non-ideal community" where crime is addressed through punishment, the role of justice is to see "deeply" into the matter. And what the just person sees, I maintain, is several things. First, the just person would recognize that the criminal is precisely the person who lacks a conscience (at least with regard to certain actions), and also that punishment is incapable of producing a conscience in the criminal. Punishment is ineffective.

Second, the just person would also recognize that resentment within the community is as much of a threat to order and stability as crime. Punishment had to be institutionalized as a practice reserved to the authorities of the community, and interpreted as a *legal* matter (rather than issue of personal injury), in order to prevent

excessive violent reprisals, mob violence, and escalating cycles of revenge. The institutionalization of punishment takes violence out of the hands of the people, where it can do the most harm, and limits its expression through legally established procedures and limits. But as such, the practice of punishment is nothing more than a compromise with resentment and the desire for revenge. Punishment is a means to addressing resentment, rather than crime.

Lastly, the just person would recognize that the practice of punishment is nothing more than a justification for human cruelty, violence, and the desire for revenge. From its beginnings in the creditor-debtor relationship, the practice of punishment involved questionable assumptions. One is that material losses can be repaid through suffering. From a straightforward economic perspective this makes no sense. Another person's suffering does nothing to restore one's material losses.

The other occurs when the concept of punishment is interpreted onto the individual's relationship to the community. As we have seen, in order for punishment to appear justified, the individual and community had to be interpreted as standing in a voluntary contractual relationship (as buyer and seller do), and crimes had to be interpreted as failures to repay a debt. And as we have seen, when crime is considered in the abstract, the only form of repayment the economic model can identify is suffering. But the suffering of the criminal does nothing to rectify any actual wrongs or injuries. In both cases (material debt or criminal action) the punishment of the criminal only serves to please those who enjoy cruelty and does not restore the actual losses, or harms.

Punishment is merely “justified” cruelty and an anesthetic for the victim’s suffering that does not address the actual issues of loss, harm, or wrongs done.

Thus, it seems that a community that has looked deeply into the matter of punishment would see that “justice” is only required in a community that is still struggling with resentment and the desire for revenge. Justice has always been the attempt by stronger parties to reach a compromise with the weaker parties it governs, in an effort to put an end to the ceaseless raging of resentment, by offering it regulated doses of anesthetization through cruelty. Justice is only a *need* when resentment threatens the stability of the community as much, or more as crime.

Ressentiment, however, is only felt by those who feel weak and powerless, and Nietzsche imagines a society (as a whole) that has a consciousness of its own power. This society would not be characterized by resentment, so resentment would not be a threat, and justice would not be a need. Thus, what began as a compromise with resentment – “everything can be paid off, everything must be paid off” – ends when the efforts on behalf of justice have successfully conquered the “ground of reactive feeling” and make justice no longer relevant. Justice has cancelled itself out through its own efforts and has freed the community from resentment, allowing it to see generously enough to no longer require punishment for crime.

But, even if we grant the existence of a community whose sense of its own power extends to all (or at least most) of its members, such that they renounce the practice of punishment, aren’t we left with at least one of the problems that we started with? Crime is

still a threat to the stability of the community (even if resentment has been overcome).

Is it at all plausible to imagine a community that leaves crime unaddressed without undermining the conditions for cooperation and eventually devolving back into fear, suspicion, anger, and resentment? If this is Nietzsche's claim, then he is being uncharacteristically optimistic and utopian.

4.5 Punishment as an Interpretation of Sanctions

The solution to this problem lies in recognizing that Nietzsche's rejection of punishment does not imply that a society should leave crime unaddressed. This requires recognizing the distinction between sanctions and the interpretation of sanctions as punishment.

As noted above, in the passage that has served as the primary inspiration for this dissertation, Nietzsche claims "we immoralists in particular are trying as hard as we can to rid the world of the concepts of guilt and punishment and cleanse psychology, history, nature, and social institutions and sanctions of those concepts..." (TI, "The Four Great Errors," 7) The key, in this case, is understanding what it would mean to "cleanse" *sanctions* of the concept of punishment. Sanctions are the community's responses to violations of the community's standards. More specifically, in the present context, they are the way that a community's legal system imposes negative responses on the person convicted of a crime.

Punishment is an interpretation of the community's already existing sanctions. The concept of punishment interprets the community's sanctions as having the aim of making

the criminal suffer, where that suffering is “earned” as a consequence for his crime. Punishment, as we have seen, has had many expectations attached to it as far as its desired outcome (deterrence, payback, reform, compensation, etc.), but the essential core of punishment has remained the same since its inception in the creditor-debtor relationship: It is “justified,” intentionally inflicted, and directly intended suffering.

A sanction, however, need not have the direct aim of intentionally causing suffering, thus a sanction need not be interpreted as punishment. We have already seen that the earliest human communities banished members who were unable, or unmotivated, to follow the rules of the community. This practice effectively removed a source of threat to the community’s stability and continued existence. I take this to be the function of a sanction (I take this to be Nietzsche’s view, as well). A sanction is meant to protect the community itself and its individual members from the threat of those individuals who cannot, or will not, follow the rules that allow for cooperation and peaceful co-existence. In early human communities, banishment (however unconsciously and unintentionally) removed threats to the community (permanently), and thus functioned as a sanction.

The practice of banishment was harsh – it was an almost certain death sentence for the “criminal.” But, however much anger was involved on the part of the banishers, however much satisfaction they felt in driving the criminal out, and however much suffering that was caused to the banished person, banishment was not understood as having the explicit and conscious purpose of causing a specific amount of suffering as a

means of compensating the community for violating the rules. The original aim of banishment was to isolate the community from a threat.

I take this distinction to be at the heart of Nietzsche's rejection of punishment.

Any community must address crime if it is to remain stable and survive, thus it will require sanctions. Nietzsche recognizes this. But punishment (as intentionally inflicted suffering) is only one possible interpretation of the meaning of sanctions. One can reject the interpretation of sanctions as punishment without rejecting the concept of sanctions for crime. Thus, Nietzsche's rejection of punishment does not imply the absurd idea that he would leave crime unaddressed and leave criminals to run wild and free.

Sanctions can be interpreted as having a different aim than causing suffering for crime. Sanctions for crime can be interpreted as having the aim of protecting the community by isolating a criminal threat. Call this the "containment model" of sanctions. Incarceration is an example of isolating and containing a threat. The criminal is locked away from the rest of the community, such that he can no longer threaten or harm others. The criminal may strongly prefer freedom to incarceration and may find the restriction of his freedom to be a source of suffering, but the aim of the containment model, as I present it, is not to cause suffering – it is only to protect the community and its members from the criminal. Thus, crime can be addressed without the aim of inflicting suffering on the criminal.

It should also be noted that Nietzsche's rejection of punishment has further reaching implications that follow from interpreting sanctions as "containment of a threat." If

incarceration is the form of sanction employed by a society, then rejecting the concept of punishment implies that the conditions of incarceration are not meant to inflict suffering, or “hard treatment,” on the criminal. As noted, the criminal may suffer from his involuntary confinement, but this is an unavoidable side-effect of incarceration, and is justified because of the criminal’s own actions. But any efforts to make the experience of incarceration unpleasant beyond this are punitive responses that produce gratuitous suffering, and are unjustified.

Nietzsche had much more to say on this matter in his earlier work, and his suggestions go even further than those I have just presented. Nietzsche claims that the criminal should be treated “as a mental patient: not, to be sure, with an arrogant show of being merciful, but with the prudence and goodwill of a physician.” Incarceration might be an important aspect of this treatment. “Perhaps he himself may find it to his advantage to live for time in custody, so as to secure protection against himself and against a burdensome *tyrannical drive* – very well!” But this is to be part of an effort to bring about the criminal’s own self-reform:

...one should neglect nothing in the effort to restore the criminal his courage and freedom of heart; one should wipe pangs of conscience from his soul as a matter of cleanliness, and indicate to him how he can make good the harm he has done perhaps to only a single person, and more than make good, through benefits he could bestow on others and perhaps on the whole community. In all of this one should show him the greatest consideration. (D 202)

Nietzsche also claims (in the same aphorism) that the criminal should be presented with an honest assessment of his chances for being “cured” (cured of his criminal drive, whether through extinction, transformation, or sublimation), and that the incurable

criminal should be offered the option of suicide, if he has become “an abomination to himself.” On the other hand, the “cured” criminal may require anonymity, a new name, and “frequent changes of residence, so that his reputation and his future life shall be as little endangered as possible.” (This suggests a “criminal protection program” to keep the criminal safe from the law-abiding community.)

Anticipating the response of a community concerned with the financial costs of housing, feeding, treating, educating, counseling, relocating, protecting, and caring for the criminal, Nietzsche counters:

...consider that the loss which society and the individual sustain through the criminal is of exactly the same kind as the loss they sustain through the invalid: the invalid propagates care and ill-humor, produces nothing, consumes what others produce, requires attendants, physicians, distractions, and lives off the time and efforts of the healthy. Nonetheless, we should nowadays describe as inhuman anyone who for this reason desired to take *revenge* on the invalid. In earlier days, to be sure, that was what one did...the invalid is in fact treated as a criminal, that is to say as a danger to the community...here the rule is: every sick person is a guilty person! And we – are we not yet ready for the opposite view? can we not yet say: every ‘guilty person’ is a sick person?(D 202)

Nietzsche also anticipates the response of the injured party. He claims that the injured person “irrespective of how this injury is to be made good, will still desire his *revenge* and will turn for it to the courts.” Anticipating the account of the creditor-debtor relationship he provides in GM, Nietzsche responds:

...and for the time being the courts continue to maintain our detestable criminal codes, with their shopkeeper’s scales and the desire to counterbalance guilt with punishment: but can we not get beyond this? What a relief it would be for the general feeling of life if, together with the belief in guilt, one also got rid of the old instinct for revenge, and even regarded it as a piece of prudence for the promotion of happiness to join Christianity in blessing one’s enemies and to do

good to those who have offended us! Let us do away with the concept of sin – and let us quickly send after it the concept punishment! May these banished monsters henceforth live somewhere other than among men, if they want to go on living at all and do not perish of disgust with themselves! (D 202)

I present this as evidence that the containment model of sanctions is not only a plausible interpretation of Nietzsche's view, but that his rejection of punishment is only the starting point for his thoughts on the way that the criminal justice system should be reformed.

Thinking through the concept and history of the practice of punishment, using Nietzsche's genealogical methodology as a starting point, provides us with both a general insight and a more specific one.

The general insight is that the critique of punishment provides us with an example of the kind of critique of morality that Nietzsche calls for (the genealogy is only a preliminary means to that critique), and is also an example of what Nietzsche means by the self-overcoming of morality. Nietzsche is directing us to examine and question the nature of the moral views and practices that we have inherited, guided by a concern for morality, rather than its outright rejection. Many long accepted moral beliefs, phenomena, and practices (such as guilt and punishment) may turn out to conceal immoral aims and motivations at their core, such that it is not simply a problem of their being put to immoral uses, but that they are in fact essentially antithetical to morality. Morality overcomes itself when it grows through directing moral critique towards its own deeply held convictions and respected practices, in an effort to reform or eliminate these in light of the concerns of morality.

Nietzsche provides an example of this when he presents the self-overcoming of justice. Justice overcomes itself when it brings about conditions in which justice is no longer a need. In the case of justice, the elimination of resentment would make it the case that the administration of justice no longer required the use of punishment as a means to make a compromise with the threat of resentment. Thus, justice would no longer be required to maintain a practice that serves to satisfy and encourage human cruelty. The end of resentment would allow a community to feel strong enough that it would no longer desire the suffering of criminals, but would use its resources towards reform, where possible, and non-punitive incarceration, where necessary.

By analogy, morality would overcome itself when it succeeded in eliminating thoughts of guilt, punishment, and resentment, and no longer confused the ethical obligations that we have to other people with aspirations to live up to an ideal of human nature.

More specifically, Nietzsche provides us with a specific insight into the nature of punishment, that will be useful for furthering the critique of other theories of punishment. According to my interpretation, Nietzsche has shown us that the practice of punishment is entirely based on the human desire to express aggression and cause suffering, and has served as more of a compromise with resentment than as any effective response to crime. As we examine theories of punishment in the next chapter one basic distinction will guide our approach. Theories of punishment can be understood as either having the aim of causing suffering or not. Theories of punishment that aim at causing suffering

always interpret the suffering produced to be towards some expected end or outcome.

Thus, those theories require a justification of both the suffering caused, and the expected results, and can fail to provide justification by either failing to provide an end that would justify the suffering caused, or by providing an end that is not attainable through causing suffering.

Moreover, Nietzsche has shown us that justice is concerned with the functioning of an entire community, which requires it to have the aim of resisting, and hopefully reducing or eliminating, resentment. Thus, any theory of punishment that fails to do this, or in fact promotes and rewards resentment, will not be serving the interests of justice.

Chapter Five: Equilibrium Through Revenge:

Hershenov's Debt Theory of Punishment

“... that humanity might be redeemed from revenge: that is for me the bridge to the highest hope and a rainbow after lasting storms.”³⁹

Preliminary: A Nietzschean Critique of Punishment

Through examination of the creditor-debtor relationship we found that punishment of debtors traces back to the community’s treatment of rule breakers by means of banishment, which traces further back to the community’s “war time” treatment of hated and defeated enemies. Defeated enemies were given no mercy. The victorious community felt free to vent any form of violence, mockery, and cruelty upon their enemies. The practice of banishment, according to Nietzsche, was simply a copy of the treatment of the community’s enemies. The customs-breaker was treated as an external enemy and expelled from the community, after which any form of violence could be vented upon him.

The creditor-debtor relationship introduces the idea that suffering can be used to pay off a debt. More importantly, in doing so, it introduces the idea that suffering is owed to the creditor, and causing the debtor to suffer is permitted by the community (or its leaders). This amounts to a primitive form of justification for causing suffering. Economic thinking moderates punishment of the debtor by making the degree of punishment commensurate with the magnitude of the debt. When the community and

³⁹ Z:II:7

individual are interpreted as standing in the creditor-debtor relationship, this thinking is also applied to the community's treatment of "criminals," such that banishment is done away with in favor of other forms of punishment that are deemed equal to the crime.

Thus, the creditor-debtor relationship functions to interpret, institutionalize, and justify pre-existing practices of cruelty. The end results which punishment is expected to achieve is a matter of interpretation and thus a matter of some will to power stamping a functional meaning on punishment. But throughout the rather numerous historical interpretations that Nietzsche presents, the means (though not the forms) of punishment have always been to cause suffering. Thus, what is essential to punishment is intentionally inflicted suffering towards some end. The underlying assumption is that human beings enjoy causing suffering, in particular when it is against a perceived source of injury. So, in this sense, punishment is *de facto* institutionalized revenge, however much that fact may be obscured by some interpreted meaning of punishment.

We have also seen Nietzsche reject one of the more popular interpretations of punishment, which is the claim that punishment is a means to awakening moral guilt. It is notable that Nietzsche decided to present and reject this interpretation, rather than other interpretations of punishment. It is my intuition that Nietzsche chose this interpretation precisely because it cuts to the moral heart of the issue. At any rate, Nietzsche's condensed discussion of this matter, taken alone, provides a wealth of resources to critique the practice of punishment.

One might look at Nietzsche's critique as primarily a practical complaint – as a matter of pointing out that a particular means is inadequate to, and perhaps detrimental to, achieving some end. And if this is the case, Nietzsche has plenty to criticize. Nietzsche points out that (by definition, it seems) the criminal lacks a conscience (at least with regard to certain actions), and that punishment cannot produce a conscience, or the pangs of guilt that accompany it. The intuitive explanation for this is that the criminal witnesses his punishers employing the same aggressive and violent means that he favors, but doing so with a good conscience. Thus, the criminal does not learn that certain actions are wrong *per se*. He only learns that one either needs to be in a position of power to perform certain actions, or that one must be more cautious about not being caught doing so.

I argued that Nietzsche's account can provide an even deeper explanation for why punishment does not awaken guilt. The feeling of guilt requires possession of a conscience (or bad conscience). The conscience is produced when a person's use of internalized aggression, in an effort to become obedient to the rules, has become their dominant instinct. The criminal lacks a conscience, which implies that his aggression has not been internalized toward obedience. Indeed, the criminal is precisely the person who directs aggression externally, towards others. The suffering that is produced through punishment is imposed by an external source. This provides the criminal with something external to resist. This not only fails to produce internalization of the criminal's aggression, it provides a means for that aggression to be used through enduring suffering.

Thus, if punishment does not internalize aggression, it cannot produce a conscience, or awaken feelings of guilt.

I take Nietzsche's critique of punishment, however, to be more than a practical complaint. I take Nietzsche's presentation of punishment in the context of the creditor-debtor relationship and early human communities to be directing our attention to the fact that human cruelty has always provided the foundation upon which punishment was built. Inasmuch as we take deliberate cruelty to others to be immoral, this seems to implicate punishment as an immoral practice.

Nietzsche's views on justice provide further support for the elimination of punishment. Justice, he claims, is a "high, clear objectivity that sees as deeply as it does generously." I have argued that to see deeply into the matter of punishment is to recognize its nature as a justification for cruelty. I also argued that seeing generously amounts to giving up the practice of punishment. But this requires resisting the demands of resentment. Justice for Nietzsche, as we have seen, is opposed to resentment and works against its destructive influence in the community.

Punishment is a compromise with resentment. Through punishment, justice seeks to provide resentment with satisfaction, while at the same time resisting its tendency to distort its targets into monsters who deserve the full brunt of resentment. Justice and resentment have different goals, and compromise with resentment seems morally objectionable. Perhaps such compromises were required to hold communities together in other time periods, but they remain compromises nonetheless. Inasmuch as justice is

opposed to resentment, it seems that its ultimate goal should be to eliminate it and bring an end to punishment.

At the very least, however, the interests of justice are not served by promoting theories of punishment that encourage and validate resentment. In what follows, I present a contemporary theory of punishment that relies on the debt model and appeals to revenge explicitly as a justification for punishment. The terms used to present and defend this theory will sound familiar after reading Nietzsche; this theory appeals to the idea of providing the victims of crime with repayment by providing them with the satisfaction of knowing that a criminal suffers. I present my own arguments and counter-examples to this theory, and show why the appeal to revenge and suffering is itself immoral and counter to several aims of morality. I conclude by proposing that the economics of suffering be eliminated in favor of the economics of “moral regard,” which requires that we abandon punishment in favor of criminal sanctions that aim ideally at the rehabilitation and restoration of the criminal, and at worst at the isolation of a criminal threat.

5.1 Introduction to Hershenov

In "Restitution and Revenge" David Hershenov defends a "debt theory" of punishment.⁴⁰ His aim is to provide a justification for the practice of punishment and in particular of the intentional infliction of suffering that is central to this practice.

Hershenov's theory explicitly connects standard economic metaphors for crime and

⁴⁰ (Hershenov, 1999)

punishment with an unorthodox appeal to the value of revenge. He uses economic metaphors to explain how crime can be understood as creating a “debt” that is “owed” to the victims⁴¹ (including the members of society at large) and how punishment can be understood as the criminal’s means of “repaying” that debt. He connects these economic metaphors with revenge by arguing that the wrongful nature of the harms and costs borne by the victims of crime produces the psychological burden of resentment in the victims. Eliminating this burden is the debt that the criminal must repay. The suffering the criminal endures through punishment is the payment made to the victims, which eliminates their resentment by giving them a feeling of “vindictive satisfaction”, which restores affairs to a state of “moral equilibrium” (sometimes, “moral symmetry”). The achievement of moral equilibrium is what ultimately justifies the practice of punishment. Hershenov also argues that punishment, understood in this way, can be seen as providing a benefit to the direct victims of crime, members of society generally, and also the criminal who is punished. Restoration of moral equilibrium benefits the victims by providing psychological compensation for the burden of resentment. By eliminating resentment it also benefits the criminal by making possible the victims’ acceptance of the criminal’s eventual return to society as a moral equal.

I argue that Hershenov's theory fails to provide a justification for punishment. I present and examine several cases that contrast possible motivations for crime with the

41 Hershenov works with an intuitive distinction between those who are victimized and harmed directly by some particular crime – the direct victims – and those who pay the costs of crime due to its general existence – the members of society as a whole. For the sake of brevity I will sometimes refer only to the “victims” of crime, which should be taken to include all those affected by crime.

motivation that is necessarily behind punishment as revenge. I argue that these cases show that Hershenov appeals to a motivation for punishment that is itself intuitively morally impermissible: causing harm to another in order to make oneself feel better. I consider objections that would claim that this motivation is morally permissible in the context of punishment. I argue that these objections fail to render the motive of revenge morally permissible. If I am correct about this, then punishment motivated by a morally impermissible motive of revenge gives the criminal warrant to resent the actions of his punishers. Thus, punishment does not eliminate resentment between the parties in question, but merely redistributes it, and this undermines the claim that punishment produces moral equilibrium, leaving punishment unjustified on Hershenov's account.

5.1.1 Preliminaries: The Failure of Retributive Theories

Hershenov briefly addresses and rejects deterrence, reform, and expressivist theories of punishment as inadequate, on the grounds that they make the "hard treatment" central to punishment gratuitous, and thus appear to have the unacceptable result of leaving incarceration unjustified.⁴²

⁴² Hershenov (like many others, I suspect) seems to assume that incarceration and the hard treatment central to punishment are inextricably linked. Presumably this is because most people would find incarceration extremely unpleasant, such that incarceration just counts as hard treatment. Thus, for Hershenov, if punishment as hard treatment is unjustifiable, then incarceration is unjustifiable. I argue that incarceration and punishment as hard treatment are conceptually distinct and need not occur together. The aim of punishment as hard treatment is to make the criminal suffer. But one might adopt a "containment model" of incarceration. Here incarceration only needs to aim at protecting the law-abiding members of society from harm, by isolating those who have proven themselves dangerous and harmful in various ways. It need not have the further aim of making the criminal's incarceration especially uncomfortable. Thus, incarceration would be justified because it protects society from further harm.

His most sustained critique, however, is directed at standard retributive theories of punishment, which he rejects on different grounds. He says that, “retribution does not seem to do much good, if any, for either the direct victim, society at large, or the criminal.” (79) A central aim of Hershenov’s project is to provide an explanation of the good that punishment does for the direct victims of crime, society as a whole, and the criminals who are punished – an explanation he takes to be a *desiderata* of any adequate justification of punishment. For Hershenov, retributive theories go wrong because they focus on the criminal and insist that the criminal deserves to be punished because of the wrongness of his action *per se*. This misplaced focus leads retributive theories to effectively ignore the harms created by crime. But for Hershenov, the reason to take crime seriously is precisely because it causes harm to the direct victims of crime and also society at large. By ignoring harm, retributive theories fail to provide an explanation for the benefit punishment provides. Moreover, focus on the wrongness of crime *per se*, leads many retributive theories to explain the function of punishment as that of “canceling” or “annulling” wrong actions through the counter-balancing right action of punishment. Hershenov, understandably, finds the notion of “annulling wrongdoing” completely opaque, and thus inadequate to justify punishment.

5.1.2 Hershenov’s Debt Theory of Punishment

Hershenov shares with retributivists the intuition that those who do evil ought to be punished. But as we’ve seen in his critique of retributivism, this will not be because of the wrongness of the crime *per se*, but rather, because of the benefit punishment provides

to the victims of crime. To explain this benefit Hershenov adopts the economic metaphors for crime and punishment used by many retributivists. These are employed to explain how crime creates a “debt” that the criminal “owes” to the victims, and which he “repays” – thereby providing a benefit to the victims – by receiving punishment.

It is easy enough to appeal to the intuition that crime creates a "debt" that must be "paid back", but to provide an adequate justification for punishment, any debt theory must provide an explanation of the economic metaphors it employs. To begin with, it must explain the nature of the "debt" incurred through criminal wrongdoing; it must identify those to whom this debt is "owed"; and it must explain exactly how punishment can provide the "currency" which constitutes repayment of the debt owed.

Hershenov explains his economic metaphors in the following way: Criminal actions produce various harms and costs to both the immediate victims of crime, and society as a whole. These harms and costs create the debt that the criminal owes.

Depending on the crime in question, there are various harms that may be produced: physical harms, including pain and suffering, injury, or death; psychological harms, including mental trauma and suffering, anxiety and fear (over one’s own safety, as well as that of loved ones), grief over losses, etc.; and financial or material harms, including loss of money, and loss of, or damage to property.

In addition to the harms done to the direct victims of crime, there are also many costs that affect society as a whole. Due to crime, law-abiding citizens must pay a higher price for their safety. Taxes are levied to employ police officers to investigate and solve

crimes, and to take criminals into custody. Taxes are also needed to support a criminal justice system that will prosecute criminals, provide for their legal defense, and incarcerate them when convicted.⁴³ Public anxiety over crime leads many to purchase security systems, property insurance, and various other protective measures that they would not otherwise need. Anxiety over crime leads many to restrict their own freedom of movement by avoiding travel in certain locations or at certain times.

These harms and costs were unfairly imposed on both the victims of crime and the other members of society. Their well-being was lowered through the criminal's actions and they deserve to have their well-being restored to its pre-crime level. The criminal's debt is to *somehow* provide compensation to the victims – to provide them with a benefit that restores their well-being (as much as that is possible).

Hershenov's theory makes (at least tacit) use of the concept of desert, but his focus is on the victims, rather than the criminal. This makes his debt theory a theory of restitution. Unlike retributivism, which is concerned to give the criminal wrongdoer the punishment *he* deserves, a restitution theory is concerned with how the criminal can repay the *victims and members of society* what *they* are owed or deserve. So, in Hershenov's theory, it is not that the criminal deserves to be punished because of the wrong done, but rather, because of the wrongful harm done them, the victims deserve to be repaid, and this is accomplished by punishing the criminal.

⁴³ The only form of judicial punishment Hershenov addresses is incarceration and he typically connects this with the topic of the criminal's eventual restoration to society as an equal citizen. This leaves his position on both capital punishment and the possibility of life imprisonment unexplained.

Hershenov's account identifies and explains the nature of the debt created by crime, the identity of the debtor, and those to whom the debt is owed. On these counts, his theory is intuitively plausible, and gives some version of what one might expect in advance from a debt theory. What is not intuitive or obvious is what explanation Hershenov (or any other debt theorist) can provide to support the claim that punishment is the currency that pays off the criminal's debt to the victims, and how that payment provides a direct benefit to the victims. This is one of Hershenov's stated goals: "My hope is that we can come to see the punishment of both those who gained from their crimes and those who did not as in some way equivalent to the repaying of a debt that makes the injured members of the public wealthier in some sense." (86)

The basic problem, as Hershenov notes, is this: "...to explain how punishment can be called a debt payment in those cases where the individual victim or injured public is not receiving any money, goods, or services from the criminal." (83) Punishment of the criminal does not provide the victims with financial restitution or compensation for their material losses. Punishment of the criminal does not provide the victims with medical treatment for their physical harms and it does not replace or resurrect the dead. Punishment does not provide psychological counseling to deal with the fear, anxiety and traumatic aftermath of victimization. Nor does punishment off-set the various costs that criminal activity creates for society as a whole; on the contrary, it adds to those costs.

In addition to this basic problem, the debt theory's explanation of punishment's benefit to the victims faces further complications. It is not the case that all criminals

benefit from their crimes, or that all victims suffer irrecoverable losses. Some crimes are merely attempted and the criminal gains nothing in the attempt. In others the stolen money or property is recovered and returned to the victims. Yet many believe as Hershenov claims, that, “Even if the criminal did not receive any unfair advantages from his crime, there is no denying . . . that most directly victimized people, and the public at large, care about a criminal’s comfort level. Their concern is in seeing it reduced by imprisonment.” (84) Thus, the debt theory will need to explain how punishment is justified in those cases where the criminal has not gained anything and the victims suffer no obvious, or lasting harms.

In other cases, where material harm is done, the criminal might have the means to compensate the victims financially for any losses. Yet here too, Hershenov believes (with many others, it seems) that financial compensation alone is insufficient to address the crime (at least in many cases). As he puts it, “There was the normal state of affairs, it was disturbed by a crime, then the criminal’s illicit gains, if any, were withdrawn or offset. But this does not restore the status quo, for the deep wound and surrounding scar tissue, along with the pique, petulance, and resentment felt by the victimized, remain untreated.” (87) So, the debt theory will also need to explain why punishment is still required in those cases where the criminal has the means to provide financial compensation for the victims’ harms and costs.

In all of these cases, once restitution is made, the criminals do not gain any lasting benefit, nor, if financially compensated, do the victims experience any lasting material

losses. Yet in all of these cases (depending on the precise circumstances) many have the intuition that punishment is not only appropriate, but required. To do justice to this intuition, a debt theory of punishment must explain why punishment is still appropriate, perhaps required, where there are no material losses to the victims or where compensation is possible.

Conversely, in some cases the criminal might spend the stolen money or destroy property and lack the means to provide financial compensation. Here the debt theory must explain how punishment provides non-financial compensation for irrecoverable financial losses.

It might be thought that the criminal *should* provide compensation for the victims' financial losses (where possible, and not to exclude punishment).⁴⁴ But many crimes do not, or do not primarily, involve financial losses. Crimes involving violence of some form cause physical and psychological harms, rather than material losses. The complication here, is for the debt theory to explain how the hard treatment central to punishment constitutes non-financial "restitution" for these non-financial harms.

To summarize: Because of the complications mentioned above, the debt theory must show that punishment provides compensation for financial losses (where financial

⁴⁴ One might also believe that financial compensation (where possible) of the victims by the criminal, for non-financial harms received, is an important feature of some punitive responses to crime, but, on Hershenov's account, financial compensation alone would not amount to punishment, and thus would not address the resentment produced by crime.

It is not clear why Hershenov's account could not take some financial damages awarded to be more than mere compensation. If the the criminal had the means, and the damages awarded were severe enough, it seems that this, by itself, might produce the suffering that is central to punishment, which would provide vindictive satisfaction to the victims.

compensation is impossible); provides compensation for non-financial harms; and is required (in some cases) in order to provide something more than financial compensation alone. Since the compensation that punishment provides is non-financial in nature, this indicates that punishment provides a benefit to the victims by compensating for something other than, or in addition to, the obvious costs and harms of crime detailed above.

Hershenov takes this additional something to be the "pique, petulance, and resentment" felt by the victims over the costs or harms they *wrongfully* incurred.

Hershenov says this:

Even if the crime were unsuccessful...the victim would still be left greatly bothered by the arrogance and contempt the criminal showed toward him, for it was this very state of mind that initially cleared the way for the villain's transgression. If the criminal had more respect for his target, he would not have been able to violate the latter's wishes through his attempts. So, even if the attempt failed, it still implied the existence of certain attitudes that would insult and enrage the intended victim. (87)

Put simply: Punishment addresses, and somehow assuages and eliminates, the victims' resentment; this is the benefit it provides. First there was the normal state of affairs, where no resentment existed between the individuals in question. Then a crime was committed, which produced *not only* a harm of some type, *but also* resentment towards the criminal over that harm. Resentment affects the emotional health of the victim in a burdensome and unfair way, and threatens to undermine it, if left unaddressed.

Punishment of the criminal allows resentment to be expressed, spent, and satisfied, thus restoring emotional health to the victim.⁴⁵

According to Hershenov, punishment can then be understood as the “equivalent of repaying a debt that makes the injured members of the public wealthier in some sense.” (86) The burden of resentment is the debt that the criminal pays off through the currency of punishment. Punishment eliminates (as far as is possible) the resentment of the victims, and thereby provides them with the benefit of restored emotional health.

But how could the suffering of the criminal address the victims’ resentment and restore their emotional health? Hershenov claims that punishment accomplishes this by offering the victims a very basic “benefit”, one that pre-dates the law and organized civic life; a “benefit” identified and critiqued by Nietzsche in the “Second Treatise” of *On the Genealogy of Morality*. This benefit is the satisfaction of taking revenge – the pleasure

⁴⁵ Some readers might notice similarity between Hershenov’s view and Herbert Morris’ work on punishment. Morris also employs debt metaphors to explain crime and punishment. For Morris, crime produces a psychological imbalance between the criminal and the law-abiding members of society. Those who follow the law restrain themselves from certain actions, with the expectation that others will do the same. Restraining oneself is a psychological burden that the law-abiding citizen takes on as the cost of living with others in a mutually beneficial social arrangement. The criminal who breaks the law throws off the burden of self-restraint, and gains an unfair advantage over others. The criminal gains from this – even if the gain is only the feeling of having thrown off that burden for a time. This burden is the debt the criminal owes to the other members of society. Punishment balances things, by imposing the burden of suffering on the criminal to replace the burden of self-restraint he avoided by committing a crime.

Hershenov finds Morris’ view implausible. For one, he finds Morris’ view of the burden of self-restraint overblown. Criminals are the exception, rather than the norm. For most people, serious criminal acts are not even considered, thus the need for self restraint is minimal at best. Thus, self-restraint is not a burden most people bear, that can be added to by the criminal’s actions. Moreover, Hershenov claims that Morris’ does not satisfactorily explain how the direct victims of crime are made better off by the criminal’s punishment. If punishment is a burden placed on the criminal to replace the burden of self-restraint he unfairly threw off, this still leaves the victims worse off than they were prior to the crime. This overlooks both the real victims of crime and the actual harms and costs it produces. Hershenov’s attempt to address these issues will be made clear in the body of this paper. For Morris’ view see: (Morris, 1968, 1971, 1981)

we feel when we strike back and cause suffering to those who have caused us to suffer. Thus, it seems, the victims' resentment over wrongful harms and costs fades as it is replaced with the satisfaction of knowing that the criminal has been made to suffer.

Hershenov recognizes that joining the concept of revenge with the concept of restitution is at odds with a great deal of moral and legal theory. Many moral and legal theorists would maintain that justice and revenge are antithetical. But, Hershenov claims that it is reasonable to question this assumption, since, on alternative theories of punishment, no satisfying benefit can be provided to the victims of criminal wrongdoing.

Hershenov's appeal to revenge builds on the reactive attitudes that are part of our normal responses to wrongdoing. Crime is a wrong that disrupts the normal, everyday lives of those affected by it. The criminal's actions – even in the case of unsuccessful, attempted crimes – manifest disregard for the rights and concerns of others. This disregard produces anger, indignation, insult and resentment. These reactive emotions target the criminal and seek expression. This much is consistent with a great deal of contemporary moral psychology. Hershenov goes further, though, by appealing to the primitive human need (however much we, or he, might wish it otherwise) to retaliate against those who cause us harm, especially when that harm is brought about by an action that manifests disrespect, disregard, or contempt for our moral standing. Taking revenge

gives us a primitive sense of satisfaction when we are able to inflict suffering on those who have wrongfully caused us to suffer.⁴⁶

Punishment, then, is an expression of a primitive need for revenge, and the satisfaction it provides can be the payment made by the criminal to the victim.

Hershenov states, “The vindictive pleasures accompanying legal punishment can either increase the value of any other form of payment received, or, where the criminal is destitute, can actually take the place of receiving financial compensation.” (80) The criminal’s suffering counts as payment (in part, or full) because the victims take pleasure in it, which relieves them of the burden of resentment, making them better off than they would be absent punishment.⁴⁷

But it is important to note that, on Hershenov’s account, it is not the “vindictive satisfaction” of revenge *itself* that justifies punishment, but what it produces. Revenge plays a role in his theory because it re-creates (at least approximately) the pre-crime equality that existed between the criminal and victims. Hershenov understands the desire

⁴⁶ Hershenov makes a distinction between what he terms “judicial vindictiveness” and “unjust” vindictiveness. He describes “judicial vindictiveness” as motivated by an egalitarian sentiment that aims at restoring “moral symmetry” between those who have been divided by an injury. He describes “unjust vindictiveness” as “a sadistic delight...experienced when a man is punished more severely than he should be and / or his deprivation is not accompanied by a recognition of his right to someday be restored to civil society as an equal.” For Hershenov, the existence of judicial vindictiveness shows that the aggrieved victims do not wish to make the criminal’s incarceration as unbearable as possible, and recognize the criminal’s right to eventually be restored to society, hopefully after experiencing remorse and recognizing punishment as appropriate.

⁴⁷ The claim that punishment eliminates resentment in the victims (as much as is possible), and thereby restores moral equilibrium, can also be seen as providing a benefit to the criminal. Once the criminal has paid his “debt” through punishment, resentment is eliminated and this is what makes it possible for the victims to accept the criminal’s restoration to society as an equal. This is the “atonement” aspect of punishment for Hershenov. If the criminal comes to see his punishment as appropriate, he may also see his suffering as a way of atoning for the injury and wrong done to the victims.

for revenge (through punishment) as motivated by an egalitarian sentiment that seeks to re-establish moral equilibrium (or symmetry). He claims that, “When the criminal is lowered to the state the victim was at before the latter’s compensation, an equality is restored between the criminal and the victim. Both have suffered equivalent harms.” (88) Thus, the victims’ desire for revenge includes a desire to “get even” with the criminal. Punishment achieves this by lowering the well-being of the criminal (through suffering), and raising the well-being of the victims (through vindictive satisfaction), to something like their pre-crime status. (87) After punishment has been carried out, both the victims and the criminal have suffered commensurate harms. The victims feel both vindictive satisfaction and the sense that they are now “even” with the criminal, this eliminates (or at least, mitigates) their resentment, and moral equilibrium has been restored. It is this moral equilibrium that punishment aims at, according to Hershenov, which ultimately justifies the practice of punishment.

Since the ultimate purpose of punishment is to restore moral equilibrium, and this is understood as the elimination of the victims’ resentment, Hershenov’s theory draws no distinction between punishment and restitution. They are one and the same. The criminal’s punishment-induced suffering is able to count as payment to the victims because they take pleasure in the fact that the criminal suffers. This pleasure in the criminal’s suffering satisfies their desire to “get even” and eliminates their resentment. The criminal’s suffering just is the payment made to the victims as restitution for the unfair burden of resentment over the harms and costs of his crime.

5.2 Critique of Hershenov

The ultimate justification for punishment, according to Hershenov, is that it restores moral equilibrium. Crime leaves the wound of resentment and punishment is the treatment for this wound. Punishment deliberately lowers the criminal to the state the victim was in subsequent to the crime; he suffers a commensurate harm. This provides the victims with a sense of satisfaction that eliminates their resentment. By eliminating resentment punishment returns a situation unbalanced by crime to its pre-crime status (as nearly as possible). Somehow this amounts to “moral” equilibrium.

I aim to question the claim that punishment restores moral equilibrium in the manner just described. There are reasons to doubt this claim that arise when we compare the intentions that *might* motivate a crime with the intentions that *must* be present when resentment and revenge are the motivations for punishment. On Hershenov’s account, moral equilibrium is achieved when resentment between two parties is eliminated. Punishment is supposed to accomplish this. I, however, believe that punishment is more likely to further undermine moral equilibrium than to restore it. My thesis is this: While various motives for crime warrant resentment of the criminal by the victims, it seems that the motive behind punishment – the desire to make oneself feel better by causing another person harm and suffering – warrants just as much, if not more, resentment of the victims by the criminal. Thus, punishment does not eliminate resentment between the parties in question, it merely redistributes it, and does not produce moral equilibrium.

5.3 Crime and Motivation

A criminal is someone who violates the law, and thereby violates the rights of others. In some cases this might be due to ignorance of the law. Commonly enough, it can also be due to a willingness, on the part of the criminal, to disregard the rights of others in order to pursue satisfaction of personal interests. At the extreme, the criminal could be motivated to violate the rights of others out of an active desire to cause harm, taking some satisfaction in doing so. In any case, and to varying degrees, the criminal is a potential threat to the well-being of others (and to overall social stability), out of ignorance, disregard, or malice with regard to their rights.

But while it can be granted that any crime is a potential threat to society and its members, it must also be noted that consideration of the motive for a crime plays an important role in justifying how we respond to the criminal. In our judicial system, sentencing is often, and it seems, should be, sensitive to the actual motive behind a crime (as far as that is possible to know with any assurance). We judge that some motives justify more resentment, and harsher sentences, than others, even with regard to crimes that have the same consequences. If motives matter to us – if they make a difference in our judgments about resentment and punishment – then it seems especially important to take them into consideration when assessing a theory of punishment that is so tightly linked to resentment and its elimination.

In what follows, I present a series of cases that differ only with respect to the motive behind the crime and contrast these with the motive for punishment that Hershenov

endorses: A desire for revenge (or as I've put it, the desire to make oneself feel better by causing another person harm and suffering).⁴⁸ Assume in each of these cases that we have a criminal who has been justly apprehended, tried, convicted, and sentenced.

According to Hershenov, the sentence that is carried out must aim to cause the criminal to suffer; it must lower the criminal's well-being, in order to satisfy the victims and thereby raise their well-being.

5.3.1 Case 1: The Malicious Driver

Imagine a person who has spent too many years in Los Angeles traffic. As a commuter he has endured countless delays directly due to the unconscionably slow, self-entitled saunter of the average pedestrian. His anger and hatred towards these lazy beings grows, until he decides to do something about it. He sets out one morning in a cold rage, speeding through a residential area, looking for a victim. He spies one dallying with a smart-phone in a crosswalk and runs him down, killing him instantly. When apprehended

⁴⁸ By glossing the motive of revenge as "the desire to make oneself feel better by causing another person harm and suffering" I am not claiming that victims motivated by revenge see their actions under that description. Some will be retributivists who believe that the criminal deserves punishment, and see their actions as "pursuing justice". It might be important to them that the sentence is carried out by the proper authorities and thus counts as an instance of "justice being done". Others might admit that they simply desire the criminal's suffering and will enjoy the fact that the criminal is made to suffer as they were. They might be equally (perhaps more) satisfied with personal vengeance, if that were the only way to make the criminal suffer.

Nevertheless, the desire in each case, is to cause the criminal suffering, and only that will satisfy this desire. If the criminal is not punished, or not punished sufficiently, or somehow takes pleasure in what was intended as punishment, then the victims' desires will be thwarted. If these desires are thwarted, then the victims experience displeasure. Thus, when they act, they aim at bringing about that which will satisfy their desires. Satisfaction of this desire brings them pleasure. Thus, they aim to make themselves feel better by satisfying the desire to cause another person to suffer.

he is still chortling with unmitigated glee over the stunned look on his victim's face, just before impact.

This driver set out with the explicit goal of doing harm to another and succeeded. The actions he took were designed to bring about harm. He was motivated by anger, hatred, and the desire to inflict harm on a source of his emotional suffering. He is guilty of murder.

There are a number of harms and victims in this case.⁴⁹ One person was killed.

Others – his family, friends, and anyone who depended on him in some way – experienced various psychological harms from this loss, perhaps financial harms as well.

There was a “benefit” of some sort to the Malicious Driver. He experienced the satisfaction of acting on, and venting, his anger and hatred, and achieved some relief from the burden of the same.

It is easy to imagine the surviving victims feeling a great deal of resentment over the harms done to them through this criminal's actions, and rightly so. He had a desire to express his anger, hatred, and frustration by harming others; he acted on that desire and took pleasure in satisfying it. Now there is resentment where there was none before, resulting in moral disequilibrium. The victims' resentment targets the criminal and they desire to “get even” with him. According to Hershenov, this desire can be satisfied when

⁴⁹ In a footnote, Hershenov defends the idea of non-experiential and posthumous harms and benefits, and the possibility that punishment could be understood as providing such a benefit to the deceased victim of a crime. I do not intend to either challenge or defend the notion of non-experiential or posthumous harms and benefits in this paper. I treat the “direct victims” of crime as both those who were victimized and harmed “in person”, and those who were harmed because of their close relation to that victim. Thus, in the cases presented that involve death, I primarily have in mind the surviving loved ones of the deceased, who may feel resentment and a desire for revenge against the criminal.

the criminal is made to suffer as well. When he is punished, the lowering of his well-being gives the victims a feeling of vindictive satisfaction, which assuages their resentment and thereby raises their own well-being. Moral equilibrium is achieved. If this criminal were not punished, the victims' resentment would remain, and this psychological burden would not allow them to tolerate the criminal's continued presence in society. Disequilibrium would remain.

Here we seem to have a case where the motives and actions of the criminal are symmetrical with the motives and actions of the victims who seek his punishment. One might say that the criminal acted so as to relieve himself of the psychological burden of unexpressed anger and hatred by causing harm to another. He took satisfaction of his own desire to count more than the victim's rights and well-being. In response, the victims treat the criminal in the same way that he has treated them. They relieve themselves of the psychological burden of unexpressed resentment by causing the criminal to suffer harm through punishment. Once the burden of resentment is removed, moral equilibrium is said to be restored. The symmetry between the actions and motives of the criminal and the victims in this case seems to perfectly illustrate the dynamic of resentment, revenge, and restoration of moral equilibrium that Hershenov envisions.

I believe that there are problems with this case, but I will be in a better position to show what those problems are after having examined cases where the motivational symmetry seen here is lacking. With that goal in mind I will continue on to the next cases before returning to the "Malicious Driver".

5.3.2 Case 2: The Thoughtless Driver

Imagine a college student speeding through a residential area to avoid being late for class. As he does so, he has no thought at all about the possibility of pedestrians crossing his way, nor that his speed is too fast to safely stop should he encounter one. He is completely consumed with the thought of arriving to class on time. He sees a pedestrian in the crosswalk and swerves to avoid him, but is unsuccessful. The pedestrian is struck and killed instantly.

The “Thoughtless Driver” did not set out to harm anyone, but in fact did. When he acted, he did not stop to consider how his actions might harm others, and that oversight led to a fatality. This driver was certainly at fault and blameworthy for this thoughtless inattention to the possible consequences of his actions, and his crime seems to fall into the category of negligence. If he was capable of considering the possible consequences, then he should have done so, but did not. And if he was incapable of doing so (for whatever reason), then he had no business driving in the first place, but did. Thus, it seems, arguably, that the victims are warranted in resenting his negligent actions.

This case involves the same harms seen in the last one. A person was killed, and those connected to that person suffer in various ways from that loss. There seem to be no benefits from this crime (I’m assuming the driver never made it to class), so there are no illicit gains to be returned or off-set.

Assume that the surviving victims feel resentment over the harms done to them by this criminal’s thoughtless action. Now there is resentment where there was none before,

resulting in moral disequilibrium. The victims' resentment targets the criminal and they desire vindictive satisfaction. When the criminal is punished, and suffers like them, they will have it. Suffering lowers his well-being, which pleases the victims, assuages their resentment, and thereby raises their well-being. Moral equilibrium is restored.

On closer examination, however, it seems that moral equilibrium is not restored. It seems that punishing *this* criminal to satisfy the victims' desire for revenge produces greater disequilibrium than was produced by the crime itself. Note that this particular criminal did not desire or intend to cause harm, or even dismiss the possibility of causing harm as having less importance than the pursuit of his own interests. He simply failed to think about the possible consequences of his action. However, on Hershenov's account, the victims who seek his punishment do desire, and intend, to harm the criminal. The criminal must be harmed and made to suffer in order to provide them with the vindictive satisfaction that will address and eliminate their resentment and restore moral equilibrium, which is their aim. And, it is not only the case that the victims *in fact* desire the criminal's harm, but they *must* desire his harm for Hershenov's account to work. If the victims do not have the desire that the criminal be harmed, then there is nothing for punishment to satisfy. If punishment does not provide vindictive satisfaction to the victims, then it does not provide them with a benefit that eliminates their resentment, and moral equilibrium is not restored.

It seems plausible to suggest that the victims are thus guilty of (and are encouraged by Hershenov's view to be guilty of) a more morally dubious motivation than the

criminal had in acting as he did. The criminal had the motive of getting to class on time. The victims have the motive of causing another person to suffer. Of these two motives, the motive of causing the suffering of another, involves more disregard for the well-being of the other and warrants more resentment than mere thoughtlessness. Thus, if this criminal is punished from a motive of revenge, he would be warranted in resenting his punishers, so punishment has not eliminated resentment, it has merely reproduced it (and more of it) elsewhere, and moral disequilibrium remains.

An obvious objection to this case is that it does not involve disregard for the rights and well-being of others on the part of the criminal – it does not illustrate ill-will – and so does not present us with a case of clearly justified moral resentment. As a case of negligence, it involves too much ignorance or stupidity on the part of the criminal (however culpable). Even the victims might recognize this and (with the thought “there but for the grace of God go I”) take resentment to be unjustified, thereby dissolving their desire for revenge. If this is so, we need a case where the motive for crime clearly justifies moral resentment, such that punishment can address resentment in the way Hershenov maintains, making punishment appropriate and justified on his terms. To address this issue I will consider a case that involves conscious disregard for the rights and well-being of others.

5.3.3 Case 3: The Risk-Taking Driver

Imagine a salesman speeding through a residential area in order to make an important business meeting that will net him a much needed commission and bonus. Imagine too,

that he is fully aware that his actions risk harm to others, but that he discounts the likelihood of causing harm through various rationalizations – “there shouldn’t be many pedestrians at this hour”, “I’m a very experienced driver”, etc. – in the hope that everything will work out in his favor. He does not want to cause harm, but believes that the satisfaction of his own ends makes risking harm worthwhile. He does not believe that making his meeting is important enough to warrant causing actual harm, so when he sees the pedestrian he swerves to avoid him, but is unsuccessful. The pedestrian is struck and killed.

The case of the Risk-taking Driver has several features in common with the Thoughtless Driver. Neither of these drivers set out to cause harm. Both attempted to avoid causing harm when it became imminent. The crucial difference with the Risk-taking Driver is that he was fully aware that his actions could cause harm, but took his own ends to make taking that risk worthwhile. Because of this awareness we can say that this driver consciously intended to drive dangerously, consciously put others at risk, and consciously disregarded their rights and well-being, whereas the Thoughtless Driver did not see his actions under that particular description.

The Thoughtless Driver put others in danger and caused harm through his ignorance or stupidity. The resentment that others feel toward him might be tempered with pity, in some instances, or with disgust or contempt in others. The Risk-taking Driver, however, knew better, yet still acted in the same way. He recognized the possibility of causing harm, but chose to disregard this as a reason for driving safely. For that reason, it seem

plausible to claim that the Risk-taking Driver is more blameworthy than the Thoughtless Driver. It also seems plausible that, if they are made aware of his disregard, and see him as more blameworthy, the victims will be warranted in feeling more resentment toward him than they would towards the Thoughtless Driver, and will take him to be a more apt target of their desire for vindictive satisfaction. Thus, the case of the Risk-taking Driver produces more, and more obviously warranted, moral resentment, and results in more, and more obviously genuine, moral disequilibrium.

But, despite the different intentions of the Risk-taking Driver, punishment still appears to create more moral disequilibrium than the original crime. Though this driver recognized the possibility of causing harm, he did not set out to cause harm and preferred to avoid it. If, counterfactually, the pedestrian had been walking on the sidewalk, rather than crossing the street, he would not have altered his plans in order to strike him. This, and the fact that in the actual case he attempted to avoid striking the pedestrian, shows that this driver never had the aim of causing harm. He only intended to get to an important meeting on time (while, admittedly, consciously taking the risk of causing harm). Causing harm was not constitutive of his desire to arrive on time, thus he cannot be accused of intending to cause harm in order to satisfy his desire.

The victims who desire that he be punished, however, must have the intention to cause him harm. Of the two motives present, the victims' motive seems, again, more morally questionable than the driver's motive, and seems to warrant more resentment. If the driver is aware that the victims intend to harm him in order to eliminate their

resentment and make themselves feel better, he will be warranted in resenting his punishers. So, if the victims punish the driver, motivated by their desire for revenge, punishment will not eliminate resentment, it will simply redistribute it (also, perhaps, produce greater amounts of it) and leave the disequilibrium unresolved.

Objections can be raised about this case as well. Like the case of the Thoughtless Driver, the case of the Risk-taking Driver does not present an agent who is willing to harm others. Whereas ignorance and stupidity played too much of a role in the first case, it might be thought that over-confidence, a poor assessment of his own driving skills, or an over-reliance on luck played too much of a role in the second case. Downplaying the possibility of causing harm (however foolishly), because of a preoccupation with one's own ends, does not amount to ill-will, and thus, might not warrant the genuine moral resentment that Hershenov's theory supposes. To address this problem, the next case will involve an agent who is willing to harm others in order to achieve his own ends.

5.3.4 Case 4: The Getaway Driver

Imagine the getaway driver for a bank robbery, speeding through a residential area in order to elude police pursuit. This driver does not have the goal of causing harm – all things considered, he would rather avoid doing so – but he is strongly committed to avoiding capture, and so will not hesitate to inflict harm if deemed necessary to that end. He sees a pedestrian crossing the street, but knows that slowing to avoid him will allow the police time to block his path, so he runs the pedestrian down, killing him.

The Getaway Driver, like the Thoughtless Driver and Risk-taking Driver did not set out with the explicit goal of causing harm. But unlike them, he was not only aware of the possibility of causing harm, but was also aware that, in certain circumstances, he would willingly cause harm where necessary to achieving his own ends. When he caused harm, he did so willingly, and as a result of his own intentional actions. He knew what he was doing, and that he wanted to do it. Here it seems plausible to claim that, if the victims are made aware of his willingness to cause harm, they would judge this criminal more blameworthy for his actions, more deserving of resentment, and a more apt target of their desire for vindictive satisfaction, than either the Thoughtless Driver or Risk-taking Driver.

The case of the Getaway Driver appears to involve obviously warranted moral resentment, and results in obvious moral disequilibrium. The question now is whether the intentions and actions of the criminal in this case are symmetrical with the intentions and actions of the victims who desire his punishment.

A defender of Hershenov's view might argue that the symmetry here is obvious. This criminal was clearly willing to harm others when they stood in the way of his own ends. The victims, it seems, are justified in resenting the quality of his will. Their resentment fuels a (perhaps) natural desire to repay this criminal's willingness to harm others in kind. Thus, the victims' intentions to harm the criminal appear to mirror the criminal's own intentions to harm the victims. They wish to repay intentional harm with intentional harm (and have the criminal repay them with his suffering). If the criminal's ill-will is

not addressed through punishment, their resentment will remain undiminished.

Punishment, on Hershenov's account, will placate their legitimate resentment and restore moral equilibrium.

But, on closer examination, at least some asymmetry exists between the intentions of the criminal and the victims who desire his punishment. While the criminal wrongly took achieving his own ends to warrant causing harm to others, harming others was not the fundamental aim of his action. More importantly, he preferred to avoid causing harm, and would have done so, if harm seemed unnecessary. If, counterfactually, the pedestrian had been on the sidewalk, and not in the criminal's escape route, then the criminal would not have altered his route in order to strike the pedestrian (even if striking the pedestrian in this scenario would not threaten his possibility of escape). Or, if the pedestrian had been struck by the criminal, but somehow survived unscathed, the criminal would not have felt disappointment, since causing harm to the pedestrian was not constitutive of achieving his goal of avoiding capture.

The victims, on the other hand, appear to take harming the criminal to be their end. They prefer that he be harmed, and will (at least in some instances) alter their actions to insure that he does not avoid harm. If the criminal were to escape custody, the victims would encourage actions that would lead to his re-capture and eventual punishment. If the criminal were to attempt a plea bargain in order to avoid punishment, the victims would protest and endorse the rejection of that plea bargain, so that the criminal receives his punishment. If they thought that the judge tended towards leniency in sentencing,

they might give testimony to their own suffering in order to sway the judge towards a harsher sentence. If the prisoner were up for early parole, they might argue against this at the hearing. They clearly desire that the criminal suffers.

This is worrisome. Aiming to harm another (and taking satisfaction in seeing harm done) seems clearly morally worse than a willingness to harm another in pursuit of some end. This makes the motivation of the victims morally worse than the motivation of the Getaway Driver. Thus, if the victims punish the Getaway Driver out of a motive of revenge, it seems that he has grounds to resent their actions, and, once again, punishment would not eliminate, but rather, would produce more resentment, and more disequilibrium.

Hershenov, of course, could reject this. He could claim that the victims do not desire harm as an end; rather, they see harm as a necessary means to restoring moral equilibrium, which is their actual end. Thus, the intentions of the victims and the Getaway Driver are in fact symmetrical. The Getaway Driver was willing to harm others in order to achieve his own ends. The victims are willing to harm him as a means to their end of restoring moral equilibrium. When the victims alter their actions to ensure that the criminal does not avoid punishment, this is not the pursuit of harm as an end, it is the pursuit of moral equilibrium as an end. Hershenov might claim that the victims prefer not to cause harm, and would avoid doing so if it was unnecessary. However, given the situation, harming the criminal is the only way to appease their desire for vindictive

satisfaction, and thereby assuage their resentment, so it is necessary for achieving moral equilibrium.⁵⁰

⁵⁰ One could also object to this strategy by pointing out that punishment is not necessary to eliminate resentment and restore moral equilibrium. There are other ways of dealing with the criminal that might accomplish this just as effectively.

One possibility is education aimed at the criminal's moral reform. By this I do not have in mind a "moral reform theory" of punishment, but rather a program that seeks to help the criminal grow in understanding and sensitivity to moral considerations. Such a program might require the isolation of the criminal from the general population – so, might involve "incarceration" – but would dispense with the "hard treatment", that Hershenov's victims seem so interested in. In some cases, at least, this could lead the criminal to remorse, apology, and a desire to make amends (though not through suffering). If, as Hershenov claims, the victims are ultimately interested in seeing the criminal's attitude change from arrogance and disregard, to remorse, and respect for others, then it seems that this approach could eliminate their resentment and restore moral equilibrium, without the need for punishment.

Another possibility is forgiveness. In some cases, the victims might be able and willing to forgive the criminal. The central focus of Hershenov's theory are the harms caused by crime. The direct victims seem the most likely candidates to feel strong resentment towards the criminal. If they forego resentment by forgiving the criminal, there is nothing left to justify punishment: moral equilibrium is already restored.

Hershenov worries that those who are too quick to forgive are engaged in a "corrupt" form of forgiveness that is evidence of either low self-esteem (involving the victim's belief that they perhaps deserved to be treated badly), or a desire to "forget" the incident and be rid of the mental anguish produced by the crime. I doubt that either of these counts as forgiveness, but that is beside the point for Hershenov's theory. If the victim believes the harm to be somehow deserved, then it seems implausible to claim that they feel resentment. If there is no resentment, then there is no moral disequilibrium, and punishment is not justified. Similarly, if the victim simply wants to forget the crime and is successful in doing so, then it seems that resentment no longer exists, leaving punishment unjustified once again.

Hershenov claims that this is not enough to warrant foregoing punishment. Even if the direct victims of crime forgive the criminal, the other members of society have incurred the costs of crime, (which the criminal cannot repay) and so have a right to "relish" the criminal's justified suffering. Without wishing to downplay the effect that the sheer presence of crime has, even on those members of society not directly victimized, this does not seem sufficient to justify punishment in many, if not most, cases of crime. Some crimes are so terrible that they affect those not directly victimized. They create anxiety over the possibility of becoming a victim oneself. (Think of the "Son of Sam", the Beltway Snipers, or any number of school shootings.) But most crimes do not have this effect on the general populace – they occur, and are dealt with, without much awareness of their details, let alone resentment over their costs. Thus, it is not obvious how most crimes would produce enough resentment in the general populace to justify punishment, on Hershenov's account.

And, in the time-honored philosophical tradition of providing strange, sci-fi type counterexamples, one could imagine a society with some form of technology that would simply eliminate resentment in the victims, by putting something in the water, or requiring post-crime memory alteration of all victims. Resentment is easily and quickly eliminated, if it occurs at all, and moral disequilibrium is a thing of the past.

If this is a fair response on behalf of Hershenov, there is something deeply unsettling about it. It is true that this response makes the intentions of the victims and criminal symmetrical – but this doesn't salvage his account, rather, it reveals a central problem. The problem is that the intentions of the victims and criminal seem all too symmetrical, and both warrant resentment. Up till now, my examination of the asymmetrical cases (Thoughtless Driver, Risk-Taking Driver, Getaway Driver) treated the intentions of the victims who desire revenge through punishment as morally worse, and as warranting more resentment, than the intentions of the criminals. I believe this to be true, but beside the point. What matters is whether the intentions in question warrant resentment at all.

The Getaway Driver willingly inflicted harm on others in order to satisfy his own desires. He preferred to avoid the unpleasantness of prison life and killed in order to do so. This certainly makes the victims' resentment warranted. But the victims are also willing to inflict harm in order to satisfy their own desires. The victims prefer to be free from the unpleasantness of resentment and are willing to harm the criminal to do so. It seems that both parties count the well-being of others as less important than their own emotional well-being, and the satisfaction of their own desires. But if the presence of this motive provides grounds for resentment generally, then the criminal has grounds to resent his treatment at the hands of the victims, and punishment, it seems, does not eliminate resentment, it redistributes it, and this does not restore moral equilibrium.

In response to this, Hershenov might claim that a significant difference between the victims and criminal has been ignored. The criminal's actions were impermissible. He

violated the rights of others to achieve his own ends, thereby creating the original disequilibrium. The victims, however, simply want to restore equilibrium, which itself seems permissible. The harm produced by punishment is necessary to achieving that goal. Therefore, punishment is permissible, and the victims act permissibly when they punish the criminal. If the action of punishing the criminal is permissible, then the criminal is not warranted in resenting the victims, so punishment does not create further disequilibrium.

But this response begs the question. The permissibility of punishment is precisely what is at issue. Hershenov has attempted to justify punishment as permissible by claiming that the satisfaction of the victims' desires for revenge eliminates their resentment, and that the elimination of resentment produces moral equilibrium. The claim that punishment in fact restores moral equilibrium is what is meant to ultimately justify its practice. The problem I have raised for Hershenov's account is that the intention behind punishment seems itself immoral. It appears to be the same intention that the criminal acted on, which produced resentment in the victims. If the intention behind punishment is immoral, then the criminal is warranted in feeling resentment towards the victims who punish him. If punishment of the criminal warrants the criminal's resentment of the victims, then it does not eliminate resentment, it merely redistributes it, and so does not restore moral equilibrium, which undermines Hershenov's justification for punishment. Hershenov cannot address my objection – that the intentions behind punishment are immoral and undermining of moral equilibrium –

by claiming that punishment is permissible because it is a response to wrongdoing that seeks to restore moral equilibrium. That amounts to assuming that punishment is already permissible in order to show that it does not produce resentment and so does restore moral equilibrium. But this is just to claim that because punishment is permissible it restores moral equilibrium, which makes it permissible.

5.4 Eye for Eye: Immorality for Immorality

One possibility to consider is that Hershenov would simply accept the claim that the intentions behind punishment are immoral, and that this is precisely the sense of moral equilibrium that he endorses. Since the criminal acted immorally when he violated the rights of others and caused harm, the victims should be allowed to act immorally in return, by causing the criminal harm. The criminal's immoral action produced resentment in the victims, and the victims' immoral action will produce resentment in the criminal, restoring moral equilibrium. On this view, the fact that the victims' intentions are immoral and warrant resentment is a necessary feature of punishment. Producing resentment in the criminal is precisely the point. The criminal's resentment of his treatment does not undermine restoration of moral equilibrium, but rather, is part of its achievement, and punishment has its justification.

If this is, in fact, Hershenov's view, then the justification he provides for punishment is troubling, for several reasons. On its own, this account is troubling since the central concept that is meant to provide justification for punishment – restoration of moral equilibrium – seems both strained and arbitrary. Strained because the state of affairs

produced through punishment hardly seems moral. To address one wrong done (the crime) by allowing another wrong (revenge through punishment), intuitively speaking, seems to make the situation more *immoral*. This is an odd state of affairs to call “moral” equilibrium.

It also seems arbitrary. On the one hand, resentment of the criminal’s motive – a willingness to harm others in order to satisfy his own desires – produces disequilibrium between the criminal and his victims. On the other hand, resentment of the victims’ motives – willingness to harm the criminal in order to satisfy their desire to be free of resentment – does not produce disequilibrium, but rather, somehow restores it. But it is hard to see how this can be so. If the presence of resentment amounts to disequilibrium in the former, then it should amount to disequilibrium in the latter. To say that some resentment matters, and produces disequilibrium, and some does not, is morally arbitrary. The only thing that seems to distinguish the criminal from the victims is the fact that the criminal’s actions were temporally prior to the victims’s response, but the temporal

priority of the criminal's immoral action does not suffice, by itself, to make the response to it automatically moral.⁵¹

Arbitrariness of this kind is problematic. To ignore the criminal's resentment over the immoral intentions of the punishers is to tacitly endorse the claim that moral considerations apply to some persons, but not to others. It treats the criminal as someone who – because of what he is: a criminal – has less than full standing as a moral agent, who either cannot object to certain wrongs, or cannot be wronged in certain ways.

⁵¹ It seems obvious that the mere temporal priority of one action to another does not suffice to justify the second action as permissible, even when the second action is a response to a wrong done in the first.

To illustrate: Imagine that Jimmy-Saul envies Billy-Bob's brand new, shiny, pick-up truck and scratches the length of it with his keys in the parking lot of the local Shop 'n' Save. Unfortunately for Jimmy-Saul, Billy-Bob was sitting in his truck at the time. He leaps out and pins Jimmy-Saul to the ground. Jimmy-Saul is a scrawny character and poses no physical threat to beefy Billy-Bob. His immediate reaction is to apologize profusely and offer to pay for the damage. But Billy-Bob is unsatisfied with this and proceeds to beat Jimmy-Saul senseless with the axe-handle hanging from the rack of his truck's rear window.

Jimmy-Saul's malicious vandalism was wrong, and Billy-Bob certainly had a reason to resent his actions. But, Billy-Bob's response was completely unjustified and impermissible. Thus, the mere fact that some action is a response to an initial wrong is insufficient to justify that action as permissible.

Some might believe that cases of self-defense provide an obvious counter-example to this claim. Ordinarily, it is impermissible to strike another person with the intention of doing them harm. But if one is attacked first, by a person who seems capable of, and intent on, doing one harm, it is generally thought to be permissible to use violent force to defend oneself by harming and incapacitating one's attacker. Thus, an action that would ordinarily be impermissible is made permissible because it is a direct response to a temporally prior impermissible action.

But it seems that it is not the temporal priority of the initial wrong action that justifies the otherwise impermissible response, but rather the immediacy of the threat of harm. Imagine that Mike attacks Paul on the street, and is intent on, and capable of, doing Paul serious harm. But, before he can land a punch, a flash-mob of local theater students, re-creating a scene from "The Music Man", emerges from a subway exit and sweeps Mike up into their performance, dragging him down the street, and separating him from Paul. Fifteen minutes later, Paul finds Mike trying to catch his breath on the steps of the library and proceeds to pummel him unconscious. While Paul would have been justified in using physical force to render Mike unconscious at the time of Mike's attack, it seems that he is not justified in doing so just fifteen minutes later. The only reason for this, it seems, is because Mike does not pose an immediate threat to Paul.

But for many theorists, punishment is, both conceptually and morally, something that can only be applied to genuine moral agents. Infliction of “hard treatment” only counts as punishment when the target of that treatment is a moral agent. Otherwise it is something else – training, behavioral conditioning, or cruelty. This is so conceptually, because punishment is thought to express moral condemnation, so can only apply to a being capable of recognizing the expression of moral claims and condemnation. Because of this, punishment is also thought to be morally inappropriate in cases where the person to be punished is incapable, either temporarily or permanently, of recognizing the reasons for which they are being punished.

Thus, to punish the criminal out of a motive of revenge, and ignore the criminal’s resentment of this motive, treats the criminal as less than a full moral agent, and thus as someone to whom punishment should not be applied. Inflicting punishment in this way is at least morally inconsistent, if not conceptually incoherent.

The “immorality for immorality” interpretation of moral equilibrium also creates other problems for Hershenov’s overall project. One of his explicit concerns is that the practice of punishment be consistent with treating the criminal as someone who will eventually be restored to society as an equal fellow citizen. He takes his debt theory to provide for this restoration in several ways.

Punishment, according to the debt theory, provides psychological compensation to the victims, by appeasing their resentment. It thereby equalizes affairs between them and the criminal. But Hershenov claims that the victims want more than this. They also want the

criminal to recognize his actions as wrong and to feel remorse over them. Hershenov takes this desire to be based on the victims' recognition of the criminal's right to eventually be restored to society as an equal. Punishment is meant to provide for this as well, in several ways. It allows the criminal the opportunity to see his own suffering as appropriate: "...As a necessary debt, a token offering to his victims that is also essential to his own expiation." (90) This is to say that the criminal can come to see his punishment as a form of "atonement" – a way of paying off the debt he owes the victims. With that debt paid, he can rejoin society as an equal. But to see his punishment as appropriate, the criminal must recognize that he wronged his victims. He must become responsive to their suffering and resentment, so that this can lead to remorse over his actions. If this occurs, the criminal will have at last learned to see his victims as equals who deserve respect and fair treatment. So, punishment is supposed to provide a way for the criminal to be restored as an equal, not only in the sense of having paid his debt in a currency acceptable to his victims, but also in the sense that his remorse manifests his moral development into someone who recognizes others as moral equals. This allows his victims to accept his return to society as an equal fellow citizen.

The "immorality for immorality" interpretation makes it difficult to see how this could take place. If punishment is motivated by immoral, resentment-warranting intentions, it is hard to imagine how the criminal could see his own punishment as morally appropriate. And if the criminal cannot see his punishment as morally appropriate, it becomes even more difficult to imagine how he will come to feel remorse

over his actions, and learn to see others as moral equals, making it unclear how he could ever see himself being “restored as an equal”. It seems as likely (if not more) that the criminal would recognize that his punishers use the same actions (use of, or threatened use of violent force, for example) and have the same intentions (a willingness to harm others to satisfy one’s own desires), for which he was condemned. It would not be unreasonable for the criminal to conclude that he is no worse, morally speaking, than the victims he harmed, and who now harm him. The only differences between the motives and actions of the criminal and the victims are: the fact of who acted first in the sequence of events, and the fact of where each party stood in relation to the law, when they acted, both of which are morally irrelevant in this case. The fact that the first action was wrong, and then a response followed, does not in itself, and absent other relevant facts, make the response morally right. And, where one stands in relation to the law has no necessary connection to where one stands morally. Yet Hershenov’s victims seem to treat these facts as making a significant moral difference. It would not be surprising for a criminal observing this to conclude that the issue is not becoming a better person, morally speaking, but becoming a better, smarter, criminal.

To summarize this point: Hershenov imagines that, through punishment, the criminal can: recognize the wrongness of his actions and feel remorse; find the guilt he feels over his actions eased; feel contrition and even relief when the victims are satisfied by his suffering; and find himself willing to endure punishment to pay back his debt. I find it more plausible to imagine that the criminal will: feel contempt for those who give a

moral justification for a desire to harm others; harbor a sense of superiority over those who cannot endure their own suffering in the way he must; and consider himself a stronger, more hardened, more capable person than those who have not endured the same treatment. Balancing an immoral action with another immoral action seems like a recipe designed to exacerbate resentment, not eliminate it. It does not appear particularly well-suited for restoring criminals as equal citizens.

Given the untenability of the “immorality for immorality” interpretation of Hershenov’s account, I will assume that he does not intend to justify punishment in this way. I will assume that he would defend the victims’ motivations as morally unproblematic, such that the desire to punish does not undermine moral equilibrium, leaving his justification for punishment intact. To do this, he will need to defend the claim that the victims’ desire to harm the criminal is morally permissible, which would make any resentment felt by the criminal unwarranted.

5.5 Justifiable Hatred and the Desire to Cause Harm

There is a defense of something very much like the motivation we see in Hershenov’s victims offered by Jeffrie Murphy in *Forgiveness and Mercy*. There Murphy addresses a distinction drawn by Jean Hampton between three types of hatred. (Murphy & Hampton, 1988, p. 88) Salient here is the distinction between “moral hatred”, which Hampton defends, and “malicious hatred”, which she condemns. Moral hatred, says Hampton, is an aversion to some person because of what he stands for morally, which includes a desire to triumph over that person and his cause. Moral hatred involves a willingness to

see that person harmed, if harm is necessary for overcoming that person's immoral cause. Malicious hatred, on the other hand, is the simple desire to harm some person. Its object is to diminish and hurt the person in order to gain a competitive advantage over him. Hampton condemns malicious hatred as both irrational and immoral.

Murphy, however, is not as willing to dismiss aspects of malicious hatred out of hand. He suggests that we recognize another type of hatred that involves elements found in both moral and malicious hatred. He notes that not all desires to harm others are motivated by a competitive desire to raise one's own status by bringing the other low (as Hampton claims), but might be motivated by retributive feelings instead. He calls this "retributive hatred", and argues that it is sometimes justified. (Murphy & Hampton, 1988, p. 90)

Retributive hatred is motivated by retributive feelings, which include the desire that the wrongdoer be made to suffer. In this it is unlike resentment, which Murphy sees as essentially involving protest over being wronged and self-assertion in the face of the wrongdoer's attack on one's self-esteem. Since retributive hatred goes further than resentment, and includes the desire to harm the wrongdoer, the defender of retributive hatred faces the difficulty of justifying this desire.

The argument Murphy provides to justify retributive hatred relies on this claim: If it is morally permissible intentionally to do X (under a certain description), then it is surely permissible to desire to do X (under the same description). (Murphy & Hampton, 1988, p. 94)

Applying this claim directly to retributive hatred, he reasons that: If there is any truth at all to retributive theories of punishment, then it is at least sometimes permissible to cause suffering to others in response to their wrongdoing (punishment). From this it follows that it is sometimes permissible to cause suffering to others for retributive reasons. Thus, it is sometimes permissible to *desire* to cause suffering to others for retributive reasons. Thus, retributive hatred – the desire to cause suffering to others in retaliation for wrongdoing, is justified. (Murphy & Hampton, 1988, pp. 94-95)

It seems that Murphy's notion of "retributive hatred" and Hershenov's notion of "vindictive satisfaction" are sufficiently similar that Murphy's argument can provide support for Hershenov's theory. The central feature of Murphy's "retributive hatred" is the desire to harm another as a response to wrongdoing. Though Hershenov is not defending standard retributivism (like Murphy), his victims' desires to punish the criminal do seem retributive, in the general sense of that word. Hershenov's victims are angry over harms received, but more importantly, they are resentful over being wronged. It is this resentment that motivates their desire to make the criminal suffer. Thus, according to Murphy's argument, if it is morally permissible to intentionally punish a criminal for wrongdoing, then it is morally permissible to desire to punish the criminal for wrongdoing. And since causing suffering and harm are constitutive elements of punishment, it is permissible to desire to harm the criminal.

Murphy's argument initially seems to help Hershenov's account. Recall that Hershenov needs to defend the claim that it is morally permissible for the victim to desire

to harm the criminal in order to be able to maintain that punishment restores moral equilibrium. And, he must be able to maintain that punishment restores moral equilibrium to provide a justification for punishment. If the victims' desire to harm the criminal is morally impermissible, then punishment would warrant resentment and create further moral disequilibrium, and Hershenov's justification for punishment fails. But with Murphy's argument in place, it seems that the desire to harm the criminal is morally permissible, so punishment does not warrant resentment and create further moral disequilibrium, so punishment is justified on Hershenov's account.

Though it seems initially plausible, Murphy's argument does not work – either as support for Hershenov or for Murphy's own defense of retributivism. The first problem to note is that Murphy's argument for the moral permissibility of the desire to harm rests on the assumption that punishment is at least sometimes justified. Hershenov's project, however, is to provide a justification for punishment. Thus any appeal to Murphy's argument begs the question again. If punishment were already independently justified, then Murphy's argument might be taken to provide a justification for the desire to harm others. But since the objection to Hershenov's justification for punishment is that it does not restore moral equilibrium by eliminating resentment, because the desire to harm others is morally impermissible and warrants resentment, he cannot appeal to Murphy's argument for support.

Moreover, even if we were to grant that punishment is sometimes justified, Murphy's argument does not provide justification for the desire to harm others (even for standard

retributive reasons). Recall that Murphy's argument relies on the claim that, if it is morally permissible intentionally to do X (under a certain description), then it is morally permissible to desire to do X (under that same description). There are clear counter-examples to this claim. I offer two.

5.5.1 The Twisted Paramedic

It seems relatively uncontroversial that it is sometimes morally permissible to amputate a person's limb without anesthetic, when doing so is necessary to save that person's life, and the anesthetic is unavailable. Imagine that Jones is a paramedic who has a desire to amputate a limb without anesthetic, in a life or death situation. Even if we stipulate that Jones would never attempt to bring about such a situation, and wouldn't wish this situation on any particular victim, his desire seems obviously morally wrong. Jones is not expressing the desire that he would prove capable of performing this action if the situation was thrust upon him. Nor is he pointing out that he wants to be put in this situation, rather than someone else, because his skills and fortitude make it the case that he is the most likely to perform well and save a life. He is expressing the desire to have the experience of amputating a person's limb without anesthesia, with all the pain and suffering that procedure would produce.

5.5.2 Over-Eager Self-Defense

There are at least some conditions under which it is permissible to kill in self-defense. Imagine that Smith is a law-abiding citizen who would never intentionally wrong someone. He would never attempt to provoke someone into attacking him and does not

wander dangerous neighborhoods looking for trouble, but he does have a desire to kill in self-defense. Smith's desire seems similar to Jones' desire. Smith is not expressing the desire that he would prove capable of defending himself with lethal force if necessary. Nor is he claiming that it would be better for him to be put in such a situation, rather than someone else, because he is mentally and emotionally prepared to defend himself if necessary. He is expressing the desire to be given the opportunity to legally kill someone.

Neither Smith nor Jones have mere pro-attitudes towards the actions of amputating without anesthesia or killing in self-defense as unfortunate, but necessary means to otherwise good ends. Inflicting suffering, or killing, and having the experience of doing so, are constitutive of their desires. If conditions changed – the anesthesia became available; a police officer arrived and averted the attack – their desires would remain unsatisfied. Thus, the motivating force of their desires is not to help others, or protect themselves, but to cause harm under a description that would make causing harm morally permissible. This amounts to looking for a justification to act on a morally wrong motive, and to do what would otherwise be impermissible.

These counter-examples show that Murphy is wrong. The permissibility of an action itself does not show that the desire to perform that action is morally permissible. An action might be morally permissible, while the motivation behind it is morally impermissible. And if the motivating desire behind an action is itself morally

impermissible, then it seems that this warrants resentment of the agent who acts on that desire, whether or not the action itself turns out to be permissible.

It seems to me that Hershenov's victims illustrate the same motivation as Jones and Smith. They do not have mere pro-attitudes towards the criminal's suffering as a necessary means to some otherwise good end (say the criminal's repentance or reform). That the criminal be made to suffer is constitutive of their desires. If it turned out that the criminal did not suffer – he was unexpectedly delighted with his cell, overall environment, new friends with similar interests, and regular meals – their desires would remain unsatisfied. Thus, they too appear to have a desire to cause harm, under a description that would make it morally permissible, and this too, seems to amount to looking for a justification to do what would otherwise be impermissible to do.

5.6 The Malicious Driver: The Irrelevance of Symmetrical Intentions

The upshot for Hershenov's theory can be summarized thusly: Since causing suffering and harm is constitutive of the victims' desires, those desires are themselves morally impermissible. If the victims' motives in punishing the criminal are rooted in an impermissible desire to make themselves feel better through revenge – through intentionally causing, and enjoying, harm and suffering to another person – then the criminal is warranted in resenting these motives. Thus, punishing the criminal from a motive of revenge redistributes resentment, rather than eliminating it, and moral equilibrium is not achieved through punishment. If punishment does not produce moral

equilibrium, it lacks justification, on Hershenov's account, and is impermissible.

Revenge, it seems, has no place in a theory of punishment.

If this critique is correct, then it can be applied to the first case examined: "The Malicious Driver". Recall that this driver set out with the explicit intention of causing harm in order to satisfy his own anger and hatred towards slow-moving pedestrians. Of the four cases giving examples of various motivations for crime, this was the only one that illustrated symmetry between the motives and actions of the criminal and the victims who seek his punishment. This criminal desired to relieve himself of the burden of unexpressed anger and hatred by causing harm to another. He took satisfaction of his own desires to count more than the victims' rights and well-being. In response, the victims treat the criminal in the same way they were treated. They relieve themselves of the burden of unexpressed resentment by causing the criminal to suffer harm through punishment. According to Hershenov, this eliminates their resentment and restores moral equilibrium.

Someone might be tempted to think that punishment as revenge is clearly justified in this case, since the motives and actions of the victims are symmetrical with the motives and actions of the criminal. In this case, at least, we can see that punishment motivated by revenge is simply inflicting on the criminal what the criminal inflicted on others, which restores moral equilibrium.

As I pointed out earlier, with the case of the Getaway Driver, this is mistaken. What matters is whether the intentions in question warrant resentment at all. If the desire to

make oneself feel better through causing harm to another is, in fact, an impermissible motive for action, then it seems to apply to this case as well. If it was this very motive that warranted resentment of the criminal by the victims, then it seems that this same motive should warrant resentment of the victims who punish, by the criminal. Thus, even if we grant that punishment eliminates the resentment of the victims, it still warrants resentment by the criminal, and where resentment exists, there is moral disequilibrium.

Punishment then, does not produce moral equilibrium, it merely re-distributes the resentment, and Hershenov's justification for punishment fails, even in the case of crimes motivated by malicious motives.

5.7 Jailhouse Morality and The Irrelevance of Hypocrisy

It is not difficult to imagine one very intuitive objection to this conclusion. My argument turns on the claim that the desire to make oneself feel better by causing harm to another (where causing harm is constitutive of that desire, and not merely a necessary (and justifiable) means to some otherwise good end) is morally impermissible in all cases, and so warrants resentment whenever it is the motive for action. This allows me to claim that the criminal is warranted in resenting those who punish him out of a motive of revenge, which supports my conclusion that punishment, on Hershenov's account, does not produce moral equilibrium, and so lacks justification.

But someone might think that something has gone terribly wrong with my critique, if it allows the criminal to resent being made to suffer, after doing the same to others. Surely, it might be thought, the malicious criminal has no standing to complain about

hard treatment at the hands of his victims. He saw nothing about his victims to count against satisfying his desires at their expense. He did not respect their right to be treated as moral equals, and did not care about their well-being. He clearly endorsed acting with self-serving disregard of others when he committed his crime, how can he object, without hypocrisy, to others doing so as well? His self-serving attitude is simply revealed again in his objection to punishment. Surely this is sufficient to justify the claim that the criminal has forfeited his right to feel resentment over punishment for his crime.

I am not unsympathetic to this objection. There is something compelling about the claim that the criminal's objection to punishment is hypocritical. The criminal acted with selfish disregard for others initially. Protesting punishment seems like an attempt to avoid the consequences of his actions – an option he did not provide his victims – which seems like further, unrepentant, pursuit of his own interests. There is something especially disquieting about the criminal attempting to avoid punishment by appealing to morality – a consideration he seemed content to ignore when he committed his crime.

I grant that it might be the case that some particular criminal's protest against punishment is entirely hypocritical. That criminal's motivation might be purely self-interested and his appeal to morality a crassly manipulative ploy to escape the legal consequences of his action. But, whether or not the criminal is a hypocrite is irrelevant. What matters is the motivation of those who appeal to morality to begin with: the victims.

The victims resent the fact that the criminal showed ill-will. They take the kind of treatment they endured at the hands of the criminal to warrant moral resentment. That is

to say: they regard the criminal's behavior as immoral – as something that no decent person should do to another. But, if they wish to punish the criminal out of a motive of revenge, then it seems that they intend to do to the criminal exactly what they condemn as immoral *per se*. So while the criminal *might* be guilty of self-serving hypocrisy in protesting punishment, it seems that the victims are *already guilty* of hypocrisy themselves. And since they have, in a sense, set the terms of their disagreement with the criminal, by condemning his actions as morally unacceptable, it also seems that they should be held to that same standard. Thus, if the victims wish to condemn the criminal according to the standard of morality, and wish to convey that this standard is to be respected, then they need to demonstrate that they respect this standard themselves. Otherwise they are open to the charge of hypocrisy such that it seems completely plausible to claim that the criminal is warranted in resenting the arbitrary nature of his treatment.

I take the arguments provided in this paper to be conclusive against Hershenov's theory of punishment. The central problem is his attempt to show how punishment can both satisfy the victims' desire for revenge and restore moral equilibrium by eliminating resentment. It is the connection between revenge, as the desire to harm another, and any reasonable interpretation of moral equilibrium that proves untenable. But I take this problem to be rooted in a more fundamental mistake that Hershenov makes in identifying the nature of the debt produced by crime.

The problem, as I see it, is not necessarily with Hershenov's starting point – his use of economic metaphors – but in his interpretation of the debt that is owed. His focus on the harm, suffering, and resentment of the victims is what has led his theory astray. It is true that one of the most significant reasons for being concerned with crime is the harm that crime produces – both to its direct victims and to society as a whole. The harms and costs of crime are admittedly great. But once harm and suffering is identified as the principle “cost” of crime that creates the “debt” that the criminal owes, then harm and suffering become the currency in trade, and this seems to inevitably lead to the conclusion that like must be repaid with like, and that the criminal must suffer in order to repay the victims the cost of their own suffering. This economy of pain and suffering then requires a mechanism to facilitate transactions of the currency in trade, which leads to the idea of institutional punishment as the intentional inflicting of harm and suffering. To explain what makes this whole economy turn has led Hershenov to claim that the victims receive something they value through the criminal's suffering. This necessitates an appeal to revenge to explain how the victims experience vindictive satisfaction at having their revenge against the criminal, and thus see the debt as paid.

5.8 Debt Revised: Recognition of Moral Standing

I've already shown why I think an appeal to the motive of revenge cannot provide a justification for punishment. But I would like to push this further and suggest that identifying the actual debt created by crime shows that punishment (whether motivated by revenge or some other punitive motive) is undermining of the actual aim of

resentment, as well as the criminal's ability to repay the actual debt that is owed the victims.

It is natural for individuals to feel anger over a harm received. Resentment is also natural, but resentment is only justified when the cause of the harm is a culpable agent. Thus, resentment tracks the wrongfulness of the harm, rather than the harm *per se*. Resentment protests the fact that the culpable party failed to consider, or ignored, the standing of the victim and their right not to be treated in certain ways. It is protest over a failure to give due consideration, or respect to the victim. Thus, what the victim protests, through resentment, is not primarily the harm received, but rather, the lack of recognition behind the action that produced the harm. Resentment, then, seeks recognition, on the part of the wrongdoer, of the victim's moral standing and right to be treated accordingly, as well as recognition of the way in which the wrongful action violated that right.

I want to propose that it is just this recognition of standing – a sense of consideration or respect for the other person as a genuine moral agent deserving of fair treatment – that we owe to each other, and which the criminal failed to pay his victim. This is the debt that the criminal now bears, and which he owes, first to the victims, but also to the rest of us, inasmuch as we rightly recognize the standing of moral agency as such. Repayment of this debt is what the victim really desires from the criminal.

I doubt whether punishment – as defined in any theory (though especially in a theory that connects it with a desire to retaliate and return harm in kind) – can ever foster an attitude of recognition and respect for the victim on the part of the criminal. Suffering, in

general, tends to distract us from anything other than the fact of our own suffering, and how to avoid it. The greater our suffering, the easier it is to overlook other important features of the world and our own lives, including the moral significance of our actions. Intentionally inflicted suffering only makes this worse. There is a natural human tendency that rivals both resentment and the desire to retaliate against those who have caused us harm: The tendency to favor ourselves morally – to see ourselves as always in the right – and to see our own suffering as undeserved. Suffering intentionally inflicted by others on the criminal as punishment seems likely to encourage this response and so to distract the criminal from important features of the situation, such as his own culpability for wrong, the suffering of his victims, and recognition of their standing as moral agents. If anything, it seems that, if the criminal comes to recognize the wrongness of his actions, and comes to respect others, especially his victims, it will be despite the punishment inflicted, rather than because of it.

5.9 The Sanction and Rehabilitation Model of Criminal Justice

With these doubts about the morality and effectiveness of morality presented I turn to the alternative that I take Nietzsche to be suggesting. The basic practice that cannot be eliminated is the community's response to crime. Leaving crime unaddressed is a threat to the community and also shows moral disregard for the members of the community who are harmed by crime. But the community's response to crime need not be interpreted as punishment. Rather, it can be interpreted as a sanction that has a different aim than causing the criminal to suffer.

My proposal (which I take to be a version of Nietzsche's proposal) is that the community's sanctions should have the non-punitive aim of isolating and eliminating a threat to the community.

The community has an obligation to protect its members from injuries and losses and incarceration is one response to those who intentionally cause injury and loss. The problem is to justify incarceration. The reason that incarceration posed a problem for Hershenov was that he assumed that incarceration was a form of punishment, which must involve "hard treatment" and the suffering of the criminal. His dissatisfaction with the ability of other theories of punishment to provide justification for incarceration as hard treatment, led him to the debt-revenge model.

But incarceration need not be understood as having the aim of causing suffering, rather, it can be interpreted as a non-punitive sanction that aims at isolating a threat to the community in an environment where that threat can no longer cause injuries, psychological harms, or other losses to other members of the community. The criminal may in fact suffer from the loss of freedom that comes from incarceration, but the direct aim of incarceration as a non-punitive sanction is not to cause suffering, but to prevent further harms and losses. Suffering from a loss of freedom is an indirect effect of incarceration. This, of course, implies that the conditions of incarceration should not cause the criminal suffering beyond the loss of freedom of movement. (Which implies that it will not look like most of the prisons in this country.) Clean, safe, and comfortable conditions, decent food, medical care, legal counsel, education, and adequate recreation,

provide a few starting suggestions. (Which implies that the criminal's needs will be met in ways that run counter to the resentment felt by many members of the community). But if the conditions of incarceration do not aim to cause further suffering beyond the loss of freedom, then the fact that the criminal is a threat that must be isolated from others is all the justification that is required for incarceration. It is a justification that no reasonable person could reject, including the criminal.

But the community's response to crime must, it seems, aim at much more than incarceration and isolation of criminals. In some cases the criminal may prove to be incorrigible, and such a threat to the community, that life-long incarceration is the only option. But, not all criminals are incorrigible, and it would be both financially unfeasible and morally indefensible to incarcerate every criminal for life, regardless of the crime. Yet each criminal was incarcerated in the first place, because of the crime he committed and the threat he posed to the community and its members. So, if the incarceration system is not to amount to a temporary solution that removes criminal threats for a period of time (through some artificially determined appropriate period of incarceration that is itself a remnant of economic thinking) only to return them to the community unchanged, it must also address the rehabilitation of the criminal towards the end of restoring him to the community. For this, the prison system would have to include education, counseling, therapy where needed, training, and support.

Those who object to this model of criminal justice tend to do so for two reasons. The first is resentment. Those who feel resentment will remain convinced that the criminal

deserves to suffer and does not “deserve” the “accommodations” that I have suggested. To this position I maintain that punishment as intentionally inflicted suffering is unjustified and immoral, and that the pursuit of justice should make no compromise with resentment. Moreover, I argue that this position is counter to the self-interest of the community and its members. It is a failure to recognize that most criminals will return to the community. The question is: do the members of the community prefer a criminal conditioned to suffering and hard treatment, or do they prefer a criminal who has been educated and counseled and may be ready to contribute to the community and make good for his crime?

The other objection to this model of criminal justice is the financial costs. The response to this objection is similar to the last. The cost of incarcerating all criminals for life (disregarding the moral objections) would be astronomical and not an option. Most criminals will be returned to the community, and those who have been subjected to hard treatment, without education or rehabilitation, are most likely to return to crime. The costs that the community will incur through the efforts to solve those crimes, apprehend those responsible, bring them to trial, and provide for their defense, before returning them to prison, as well as material losses and harms to the victims, will also be significant. The question is: which is the better investment – education and reform, or recidivism?

Finally, for those who have moral concerns, it seems that the containment and reform model is the most plausible alternative to bring about the criminal’s recognition of the importance of standing in relations of mutual regard with other people. If I am correct

that the aim of resentment is moral protest, and that the debt that is owed by the criminal to the victim is recognition of moral standing and apology, then punishment runs counter to this aim, and those who have concern for “moral equilibrium” should consider whether another alternative should be adopted.

Chapter Six: Undeserved Survival: Velleman on Survivor Guilt

Guilt. – Although the shrewdest judges of the witches and even the witches themselves were convinced of the guilt of witchcraft, this guilt still did not exist. This is true of all guilt.⁵²

Preliminary: Nietzsche on Guilt

Through examination of Nietzsche's account of the development of guilt we learned that the feeling of guilt originates as debt consciousness in the context of the creditor-debtor relationship. This relationship employed the practice of punishment as a means of providing repayment for debt, by offering the unpaid creditor compensation in the form of the pleasure of making the debtor suffer. This produced a psychological connection between consciousness of debt and awareness of the possibility of suffering punishment. As the creditor-debtor relationship is interpreted onto the individual's relationship to the community and to the ancestors of the community, the individual's obedience to the rules of the community, and sacrifices to the ancestors, are interpreted as debts that must be repaid. As these ideas are internalized they form a primitive sense of obligation in connection to the debts that are owed.

I argued that the economic interpretation of the community's obligation to the ancestors was particularly strained. First, the ancestors (as spiritual beings) are not present to the community to make their desires, or the satisfaction of those desires known, which produces doubt over whether the sacrifices offered are adequate. Second, the nature of the ancestors as spirits makes their desires mysterious, which creates further

⁵² GS 250

doubt as to what they require. Third, the fact that the community's misfortunes are not interpreted as the ancestor's punishment, but rather, as evidence of their reduced power, means that misfortunes do not reduce the weight of debt consciousness. Instead, misfortunes create a crisis in which the community's existence is at stake. Evidence of the ancestor's declining power is evidence of the community's potential demise, which leads to further doubt as to whether adequate compensation has been provided. Thus, the community must will increasing debt, and greater sacrifices, in order to be assured of the ancestor's power and the community's continued survival. The resulting dynamic is increasing doubt over the community's ability to provide adequate compensation for its debt, and increasing anxiety over punishment, which reinforces itself by resisting the idea that the debt can be reduced. This provides the underlying psychological dynamic which will is the basis for the feeling of guilt.

I also argued that the relationship to the ancestors, or gods, conflates ethical obligations with religious obligations. Because the community understands its survival to be dependent on the gods, following the customs that the gods have established (both ethical and ritual) becomes imperative. Ritual failures threaten to anger the gods, and thereby threaten the entire community, and thus are seen as wrongs to the community that are punishable offenses. The relationship to the gods makes the obligation to keep their customs categorical in nature. Through this conflation of the ethical (what one owes to other members of the community) and the religious (how to please the gods), the

connection between living up to religious ideals is confused with fulfilling one's ethical obligations.

When the slaves of GM II inherit the values and gods of the nobles they inherit a feeling of debt that they judge themselves inadequate to repay. The suffering that results from anxiety over unfulfillable obligations and punishment leads to resentment, and leaves them open to the priest's interpretation of their condition as guilt before God. The priest identifies the source of their suffering as their own animal natures, which gives them the pretext to redirect their resentment within, and punish themselves for their own sinful inadequacy to live up to the standards of the ascetic ideal (in this case, a holy God). This is the characteristic feeling of guilt: a pervasive sense of one's inadequacy to live up to an ideal, and anxiety over the negative responses that one's inadequacy deserves, which motivates punitive feelings towards the self as an effort to somehow compensate for one's inadequacy.

Nietzsche's genealogy of guilt reveals that its underlying basis is anxiety over negative responses of other people (including gods, or God, if they are thought of as personal in nature), which is rooted in the recognition that other people find satisfaction in causing suffering, and seek pretexts to do so. This can be straightforward and honest, as in the creditor-debtor relationship, or can be concealed by the justifications for punishment provided by morality, or even further concealed by justifications provided by religion, through which those who wish to cause suffering can use the concept of guilt to initiate and enjoy the spectacle of self-imposed suffering.

My claim is that anxiety over “relational vulnerability” is the underlying phenomenon that explains human susceptibility to feelings of guilt. The human animal has had to live in communities of some form for the entirety of its existence in order to survive. Its instincts are attuned to the expectations and feelings of the other members of its community, and it has an instinctive sense of reciprocity. The human being is sensitive to the ways in which other people can feel injury and anger, whether this is rational or not, and thus is aware of the ways in which the feelings of other people can affect its own prospects. The average human being’s (now innate) sense of reciprocity leads it to attempt to recover equilibrium with its fellow human beings.

The interpretation of human suffering as guilt has had several millennia to become firmly established in human psychology, and through its moralization, has continued to exist in non-religious contexts. Thus, human beings tend to respond to situations in which they are powerless and inadequate, and face of the prospect of negative responses from others, with the assessment that they are somehow guilty. Guilt has become a reflexive response through two influences. The first, is that it is a residuum of the instinct to maintain relationships in order to survive. To feel guilt, is to engage in self-punishment, which is to suffer, which is an attempt to show reciprocity to others. The other is based on confusing an inability to live up to some ideal, with a moral failure that warrants the negative judgments of others. This is further complicated by the fact that we tend to perceive the resentment and judgment of others as an indication that we have failed to

meet some standard, when in fact, the others are merely resentful and there is no standard to be met.

In what follows I address a case in which the resentment of others leads to a feeling of guilt that is unwarranted. I argue against the claim that “survivor guilt” is a warranted feeling and employ counter-examples to provide support for my position. While I am sympathetic to the idea that guilt is, generally speaking, a feeling of “normative defenselessness,” I claim that this definition is derived from the moral feeling of guilt, and so obscures the underlying non-moral phenomenon, which is a sense of “relational vulnerability.” Thinking of feelings that are typically misidentified as guilt (such as survivor guilt) in terms of “relational vulnerability” provides a way to understand how these feelings occur, without placing the burden of guilt, and feelings of inadequacy that are thought to warrant suffering, on those who experience them. Moreover, the idea of “relational vulnerability” provides a way of explaining why it is appropriate for a person to feel pain over being the source of pain to others, without the assumption that the negative responses of those who identify that person as a source of pain amount to an ethical judgment.

6.1 Introduction to Velleman & Survivor Guilt

Individuals who survive disasters of various kinds, where others did not, report subsequent feelings of guilt connected with the fact of their survival. This despite the fact that they do not see themselves as morally culpable for acting in questionable ways to ensure their survival, or for failing to aid in the survival of others. These survivors

appear to understand what *moral* guilt is, and also seem to recognize that they fail to fall within its ambit, yet they continue to describe their own anxious feelings as “guilty”, and claim that the source of that feeling is the sheer fact of having survived a catastrophe that others did not. This phenomenon has come to be known as “survivor guilt”. Of the questions that can be raised by philosophers about this phenomenon, two seem especially important. The first is whether the phenomenon of survivor guilt bears sufficient structural similarity in its dynamics to the more common experience of moral guilt, such that it is best understood as a species of guilt, or whether it is dissimilar enough that another term would be preferable to the use of the word *guilt* (or that it should always be rendered in scare-quotes as: survivor “guilt”). The second question is whether, as a response to surviving a tragedy where others did not, survivor guilt can be defended as rational, or appropriate.

6.1.1 The Set-Up: Moral Guilt vs. Survivor Guilt

Imagine the following scenario:

Two men, Jones and Smith, board separate commercial aircraft that each crash shortly after take-off.

On impact Jones’ plane is relatively intact, but catches fire. Fearing for his life, Jones clammers over his seat-mates, shoves children out of his way, clubs the flight attendants attempting to maintain order with his briefcase, kicks those busy deploying the emergency ramp to the asphalt, and aims for their unconscious forms to cushion his landing as he leaps from the plane. Soon after his escape the plane explodes, killing everyone but Jones.

On impact Smith’s plane is torn to shreds and Smith is hurled hundreds of feet away only to land safely on an inflatable “moon-walk” at a local school carnival. Smith finds himself completely unharmed, though confused to be surrounded by children and clowns. He is the only survivor.

After the disaster, Jones has feelings of intense anxiety connected to his survival. On reflection he realizes that he feels guilty for the way he acted to ensure his survival. He believes his actions were morally wrong. Jones feels moral guilt.

Smith also experiences feelings of intense anxiety connected to his survival. On reflection he realizes that he feels anxiety over the sheer fact that he survived, where everyone else did not. Smith is unable to shake the feeling that his survival is “undeserved”. Though he realizes that he has done nothing morally wrong – his actions did not contribute to any fatalities, he did not fail to aid anyone that he might have aided – the only name he can find for his anxious feeling is: Guilt. Smith is puzzled by the fact that he feels guilt. He recognizes that, given the circumstances, he had no opportunity to act in any way, and so could not be understood as responsible for his own survival or anyone else’s misfortune. Moreover, his feeling of guilt is not connected to the fantasy that he might have mustered super-human powers of strength, courage, or quick-wittedness, such that he could have saved others, or averted the disaster altogether had he only acted differently. He knows that he never had the chance to act at all and knows that no one could have done any differently given the circumstances. Finally, Smith’s guilt is not over any feelings of thankfulness that others died instead of him. Something like the opposite is closer to the truth. He sometimes wonders if he shouldn’t have died with everyone else.

In Jones’ case it seems appropriate to call his feeling of anxiety *guilt*. It would not be unreasonable for Jones to believe that, had he acted differently, at least some of the other passengers might have survived the disaster. But Jones does not feel survivor guilt. He does not feel guilt over the sheer fact of his own survival. He feels guilt over the way he acted in order to survive and this is best described as moral guilt.

Smith’s case, on the other hand, rules out the possibility that he feels anxiety over any possible actions or inactions (or over any (arguably) unseemly thoughts or feelings he might have in the aftermath of the disaster) and so presents us with an explicitly non-moral form of anxiety, typically called “survivor guilt”. The question is whether Smith’s

feelings are best described as “guilty” feelings – feelings that are structurally similar to certain key features of moral guilt – and whether it is appropriate for him to feel this way.

David Velleman, in “Don’t Worry, Feel Guilty”⁵³, argues that non-moral survivor guilt can be understood as a rational response to the kind of situation that Smith finds himself in. Velleman claims that it can be “rational” or “irrational” for a person to have various emotions, including survivor guilt. I take his use of these terms to be synonymous with “reasonable” or “unreasonable”, “called for” or “uncalled for”, “appropriate” or “inappropriate”, etc., and not as a claim about the rationality of the emotions generally speaking. The purpose of this paper is to explain Velleman’s understanding of survivor guilt and to critique the argument he offers for its “rationality”. I will argue that Velleman fails to establish a precise structural parallel between moral guilt and survivor guilt, and as a result, the phenomenon of survivor guilt cannot be understood as “rational” on Velleman’s terms.

6.2 Velleman’s Project

Velleman’s defense of survivor guilt as a non-moral form of guilt that is, at least sometimes, rational, involves the intuition that guilt is a family of emotions that bear some resemblance to each other. (169) He proceeds to show this resemblance by first examining the phenomenon of moral guilt.⁵⁴ Through this examination he arrives at a

⁵³ All in-text citations are taken from David Velleman’s “Don’t Worry, Feel Guilty”, *Self to Self* (Cambridge University Press, 2006), unless otherwise noted. (Velleman, 2006)

⁵⁴ Velleman addresses two forms of non-moral guilt. In one version we have those who report feeling guilt over actions that cannot be plausibly construed as morally wrong. In the other version we have those who report feeling guilt where no (relevant) actions are involved at all. Velleman examines “self-disciplinary guilt” as an example of the former and “survivor guilt” as an example of the latter. Since my only issue is with his account of survivor guilt I omit any discussion of self-disciplinary guilt.

general account of guilt that shows why it is rational for a person to feel guilt over moral wrong-doing. He then argues that the phenomenon of survivor guilt has the same structural features as moral guilt and concludes that it can, at least sometimes, be understood as rational. In what follows I first explain Velleman's account of moral guilt before turning to his argument that non-moral guilt is structurally similar to moral guilt.

6.2.1 Moral Guilt: Loss of Trust & Normative Vulnerability

Velleman's account of moral guilt can be expressed succinctly: In acting wrongly, the wrongdoer violates the normative expectations of the moral community. His actions manifest a failure to recognize, or a lack of regard for, moral norms. In so acting, the wrongdoer demonstrates a lack of goodwill and thus provides evidence to others (should the wrong be discovered) that he is not to be trusted in matters that require his recognition of, or willingness to modify his behavior in light of, moral considerations. The wrongdoer realizes that, if his wrong is discovered, he will be unable to justify himself, or his actions. He has given others good reason to mistrust him and has left himself open to negative reactions that include blame, resentment, retaliation, or punishment, and his anticipation of these reactions produces the feeling of anxiety that we know as moral guilt.

From this characterization of moral guilt, Velleman proposes a general account of moral guilt as anxiety over loss of general trust as well as the potential consequences that may follow from that loss. Velleman distinguishes between two forms of trust, which I explain below. Moral wrongdoing warrants various negative emotional responses on the

part of others, which include withholding both of these forms of trust. The anxiety felt by the wrongdoer over these possible losses is described as a feeling of “normative vulnerability” (or “normative defenselessness”). I will deal with each of these in turn.

By violating the expectations of the moral community, the wrongdoer shows that he is not to be trusted in matters that require his ability or willingness to be guided by moral considerations, which gives others a reason to lose attitudinal trust (the *feeling* of trust, for lack of a better expression). The wrongdoer thereby loses standing in the eyes of others, and is (at least temporarily) “excluded from the company of those who are recognized as persons of goodwill.” (162) If his own sense of identity includes being a member of the moral community, this response alone constitutes a significant loss to the wrongdoer.

In addition, the wrongdoer has given others reason to withhold practical trust. Practical trust is a willingness to rely on another’s goodwill; to put oneself, or the things one cares about, in another’s hands, and at risk, should that person’s will turn out to be less than good. The wrongdoer’s actions are evidence that his will is less than good and this gives others reason to withhold cooperation in any shared projects that require putting themselves, or their interests, at risk by relying on his goodwill. The importance of practical trust is clear – it is the basis on which we are able to engage in projects that require cooperative interaction. So, the potential negative consequences of wrongdoing extend beyond reactions of blame, resentment, and loss of standing in the moral community, to include warranted “retaliation” in the form of a justifiable refusal on the

part of others to cooperate with the wrongdoer on his own projects. (165) Thus, the wrongdoer's own actions (potentially) undermine the only basis on which he could hope for the success of some of his own projects.

Common to the loss of both forms of trust is the wrongdoer's normative vulnerability. Velleman describes normative vulnerability as the feeling of being somehow unjustified in relation to some normative standard (in the case of the wrongdoer this is a moral standard, though Velleman will claim that one can also be normatively vulnerable in relation to non-moral normative standards, as well). (156) Normative vulnerability involves the anxious feeling that there is nothing to be said on one's own behalf and the recognition that one is helpless to defend oneself against the negative response of others. Importantly, the wrongdoer feels helpless because he recognizes that others are warranted in their responses. The wrongdoer knows that he is "stripped of a claim or entitlement" to be spared from these negative responses. (157) So, if others feel that they are no longer able to trust the wrongdoer, he will have no right to protest this loss of trust. And, if this loss of trust leads others to opt out of any risky cooperative ventures with the wrongdoer, they are (absent other moral considerations or constraints) entitled to do so, and the wrongdoer has no standing to complain about this fact.

Velleman's general account of guilt as normative vulnerability seems to map onto descriptions of the feeling of moral guilt quite well.⁵⁵ Guilt is reported as (among other things) an anxious feeling, as apprehensive expectation of some impending-but-vague future harm. The notion of normative vulnerability seems to capture this feature of guilt. By violating moral norms, the wrongdoer knows that, should he be discovered, negative responses to the quality of his will are warranted. But what he does not know in advance are the forms these negative responses will take. In particular, he cannot know how loss of trust will impact his own projects. Thus, he cannot prepare himself, but can only wait anxiously for the worst, knowing that he will have no standing to protest any loss that he might incur.

As an explanation of moral guilt, this also makes sense of the claim that it is rational for a person to feel anxiety over the possible discovery of wrongdoing. Moral guilt need not be understood as an unconscious and irrational fear of divine punishment, or an adult's lingering (and irrational) childhood fear of losing parental love. Moral guilt is an apt response to finding oneself defenseless against the very real possibility of genuinely bad consequences. Resentment, blame, being excluded from the company of persons of

⁵⁵ Part of Velleman's project is to adapt and defend Freud's account of guilt as anxiety over the possibility of punishment. Velleman's project, however, differs from Freud's in significant ways. Freud was only interested in rendering guilt intelligible, while Velleman wants to render it "rational". Thus, Velleman has reason to distance himself from Freud's emphasis on anxiety over the possibility of punishment. Freud's analysis focusses on formative experiences in childhood development, where guilt is understood as anxiety over the loss of parental love (this love being the only thing in the child's mind, it seems, that holds parental anger and retaliation in check). This analysis is not adequate for Velleman's purpose of showing guilt to be rational. Velleman wants to make sense of guilt in cases where loss of parental love is not a real issue and where there is no plausible expectation of discovery or subsequent punishment. Velleman wants to show why it is rational for adults to feel guilt, rather than merely intelligibly explain how childhood anxieties continue to haunt us as adults. Velleman does think, however, that by explaining the anxiety of guilt as anxiety over possible negative consequences that may accompany the discovery of one's wrongdoing, his account also explains Freud's focus on anxiety over the possibility of punishment.

goodwill, loss of trust, and the possible de-railing of projects one cares about are all significant losses that any reasonable person would respond to anxiously.

6.2.2 Normative Vulnerability & Survivor Guilt

I now turn to Velleman's treatment of survivor guilt. Velleman is only interested in explicitly non-moral instances of survivor guilt. Jones is a survivor and feels guilt, but it seems more plausible to think that he feels moral guilt over the actions he took to secure his survival. Velleman is interested in cases like Smith's, where actions and attitudes⁵⁶ are not an issue. Smith feels guilt over the sheer fact of having survived some life-threatening situation where others did not. The question is whether this is ever rational.

Velleman claims that survivor guilt can be understood as a form of normative vulnerability. He claims that it has the same structural features seen in his account of moral guilt, and can be understood as anxiety over possible resentment, which is, at least sometimes, rational. There are, however, important differences between moral guilt and survivor guilt that Velleman addresses, in order to explain how they remain structurally similar.

The first, and most obvious difference (since we are dealing with a non-moral form of guilt) is that the feelings that arise cannot be due to the violation of moral norms. Thus, the normative vulnerability connected with survivor guilt will have to be due to a violation of a normative standard that has non-moral (i.e. morally *neutral*) content.

⁵⁶ Velleman wants to rule out cases where the survivor's guilt might be over feelings that (arguably) could be construed as immoral, such as being thankful that one survived instead of others. Smith's case is meant to rule this possibility out. He is not thankful that he survived, he wonders if he should not have died with the others.

Velleman's appeal to the existence of non-moral normative standards seems, in itself, uncontroversial. Such standards obviously exist. There are standards that determine what counts as a well-played sonata, a perfectly executed swan dive, a good mechanic, a properly mixed Manhattan, etc. A person can be successful or unsuccessful at meeting these standards without running afoul of morality. Fumbling through a sonata, doing a belly-flop, mistaking the timing belt for the fan belt, and using the wrong ratio of vermouth to bourbon can all be understood as failures to live up to some normative standard, but none of these failures are, in themselves, moral failures. For Velleman, the feeling of survivor guilt is in a class with these failures, and involves "a judgment or perception whose content is normative in a more general sense." (156)

The second difference is that, since we are interested in cases where a person's actions are not at issue, the anxiety felt cannot be over loss of trust. Trust can be lost, or withheld, on the basis of moral or non-moral considerations, but neither of these apply to survivors like Smith.

Unlike Jones, Smith has not provided others with evidence that he cannot be relied upon to recognize and act on moral considerations, so mistrusting him over the quality of his will would be completely misplaced (and would make this a case of moral guilt, rather than survivor guilt).

Nor are there reasons to mistrust Smith on the basis of non-moral considerations. It is possible for a person to be inept or unreliable without also being morally culpable, and we are sometimes warranted in withholding our trust on the basis of that ineptitude or

unreliability, where those qualities might put ourselves or our interests at risk. I might not loan my car to an otherwise conscientious and well-intentioned neighbor whose driving skills I judge to be less than inspiring. You might not act on my sincere and well-meaning movie recommendation, if you think that my aesthetic sensibilities are lacking. Neither of us may trust our caring but easily distracted and forgetful colleague's promise to pick us up at the airport. Facts about a person's aptitudes or general personality traits can give us reasons not to trust them with the performance of certain specific tasks. But Smith's case provides us with no evidence of this sort. There is no evidence that his aptitudes or general personality traits contributed in any way to his survival, or the catastrophe itself, so there is no reason to believe that trusting Smith would put anyone, or anything, at risk.

While the anxiety of survivor guilt cannot be over a possible loss of trust, cases where trust is lost or withheld on the basis of non-moral considerations do help us to see an important (and easily misunderstood) feature of Velleman's account of survivor guilt. The judgment that a person is inept and not to be trusted with some task can be made on the basis of non-moral normative standards (what counts as good driving skills, good taste in movies, or punctuality), and failures to live up to those standards need not be morally culpable. Velleman will explain survivor guilt in a similar way: as a non-morally non-culpable violation of a non-moral normative standard. His discussion of this can be initially misleading, since he uses terms like "desert", "unjust", "justification", and "resentment" in his discussion of survivor guilt. These are terms typically found in moral

discourse, but Velleman intends for them to be understood as having a normative content that is specifically non-moral. In the discussion of survivor guilt, these terms can be taken to reflect a person's relation to a non-moral standard that occurs through no action of his own (culpable or otherwise).

Velleman claims that survivor guilt arises because the survivor possesses a good (survival; his life) that he cannot justify to others. He finds himself in the position of having nothing to say for himself, of having no defense for his own good fortune. His survival is undeserved and he feels anxiety over the possible warranted resentment of others over his undeserved good fortune. (167-68)

This account of survivor guilt is structurally similar to the account of moral guilt. With moral guilt, the wrongdoer recognizes that others are warranted in withholding trust as a response to wrongdoing, so it is rational for him to feel anxiety over this and other possible negative consequences. Similarly, the survivor is supposed to recognize that others are warranted in resenting his good fortune, so it is rational for him to feel anxiety over the possibility of resentment.

Understanding survivor guilt as anxiety over possible resentment raises an immediate worry that reveals another difference between moral guilt and survivor guilt. It might seem more plausible to suggest that others do not resent the survivor's good fortune, but rather, envy it. Other people envy the fact that the survivor lives, rather than their loved

ones.⁵⁷ If this is so then, even if it would be rational to feel anxiety over resentment, cases of survivor guilt do not involve resentment, and Velleman's account misses the mark.

Velleman addresses this worry by suggesting that we expand our notion of resentment to include a new variation that involves envy, which he terms "envious-resentment". He claims that, since we are dealing with two distinct species of guilt (moral and non-moral), we should not rule out the possibility that we could be dealing with anxiety over two distinct species of resentment.

To support this expansion of the notion of resentment, Velleman compares anger to envy. He claims that both "rise to the level of resentment under similar conditions," which are bitterness accompanied by a sense of injustice. (168) He notes that anger can be felt over any harm received, even when that harm is accidental and non-culpable, but anger gives rise to resentment when we recognize that the harm in question was brought about through moral wrong-doing or ill-will. He suggests that envy can be understood in the same way. Envy occurs when we are pained by someone else's good fortune – even when this fortune is recognized as deserved. But, Velleman claims, envy can give way to resentment when we recognize that good fortune to be undeserved. So, while non-moral anger is modified by the recognition of wrongdoing into the more familiar moral emotion

⁵⁷ Velleman considers the claim that the victims of a catastrophe are dead and thus no longer in a position to resent the survivor of that same catastrophe as a possible objection to his view. He focusses on the possibility that third parties (I sometimes refer to them as the loved ones of the victims, sometimes just as the victims) can feel resentment on behalf of the deceased – a resentment that is (perhaps) felt even more sharply, since its proper subjects are no longer alive to feel it – and claims that it is rational for the survivor to feel anxiety over providing grounds for this form of vicarious or sympathetic resentment. (168)

of resentment, non-moral envy is modified by the recognition of undeserved good fortune into envious-resentment.

The following examples should illustrate the dynamic Velleman has in mind:

Angry-Resentment: Let's say that you are helping me move into my new apartment. A heavy piece of furniture slips out of my hands and you sustain a back injury that will lay you up for weeks. Because it seems to have been purely accidental you don't blame me, but you are angry about the injury itself, and it seems, not unreasonably so.

Then you discover that, in order to keep you laid-up and unable to interview for a job that I want for myself, I cleverly and intentionally inflicted that injury by dropping the furniture at just the right moment. Velleman would claim that your original anger is modified into angry-resentment, brought on by the recognition of the injustice of my action.

Envious-Resentment: Imagine that I envy my neighbor's wealth and the lifestyle that goes with it. He's able to do all of the things that I would like to do, but cannot: travel the world, send his children to the best schools, eat at expensive, trendy restaurants whenever he pleases, etc. I want my neighbor's wealth and lifestyle, but recognize that this is an unlikely prospect. As my envy grows, my fantasies turn to scenarios in which my neighbor suffers financial ruin and loss of his comfortable lifestyle. This is envy at its worst.

Then I discover that my neighbor is a member of the Mafia. His wealth comes from exploiting others through loan-sharking, prostitution, and the protection racket. Now

there is a new development in the feelings I have toward my neighbor. I come to see his wealth as undeserved and resent the fact that he exploits others to make his money. Velleman would say that my original envy is modified into envious-resentment – a rational, warranted, and appropriate response to an undeserved good. Initially, I only envied my neighbor for being wealthier and more successful than I, but my discovery that his wealth was gained through illicit means transformed that envy into envious-resentment. I now envy his wealth and lifestyle, but also resent the fact that these goods are undeserved.

In both of these cases the original emotion is modified by the recognition that something was undeserved. You didn't deserve my treatment of you, which modified your anger into angry-resentment. I believe that my neighbor doesn't deserve his wealth and lifestyle, which modifies my envy into envious-resentment. Velleman's claim is that "envious-resentment and angry-resentment form a natural pair of emotions embittered by a sense of injustice." (168) This sense of injustice makes the resentment warranted, which, in turn, makes anxiety over the possibility of being the target of either form of warranted resentment a rational response. When this is applied to the case of survivor guilt, we see that a survivor like Smith finds himself in possession of a good that is undeserved, and with nothing to say in his own defense, he recognizes that others would be warranted in feeling envious-resentment, making it rational for him to feel anxiety over this prospect.

6.3 Initial Worries About Envy and Resentment

There are a number of concerns that arise from Velleman's introduction of envy into his account of survivor guilt. Velleman addresses one of them. To the possible complaint that envy is never a warranted emotion, Velleman simply goes against the grain and defends it as (at least sometimes) warranted. He claims to see no reason to deny that the victims of misfortune have grounds for envying the more fortunate, or for resenting those whose good fortune is undeserved.⁵⁸ (168)

I am sympathetic to the idea that envy is *never* a warranted emotion, but will grant this claim to Velleman. My critique does not depend on settling this issue, but will instead show that Velleman's explanation of the rationality of survivor guilt fails to meet the standards set by his own general account of guilt.

Another concern with Velleman's account that can be addressed, and set aside, results from misunderstanding his view. This is the worry that any appeal to the notion of

⁵⁸ Another worry about using envy to explain survivor guilt is that Velleman's account requires a very specific understanding of envy that is not uncontroversial. Envy can be generally characterized as an emotion that involves pain or distress over the good fortune of another. But for some, the proper understanding of envy is, more precisely: pain or distress over the *deserved* good fortune of another. This definition of envy is based on understanding it as primarily about competition between a subject and a rival, where the rival's possession of some good, rather than the subject's lack of that good *per se*, is perceived by the subject to diminish his own sense of self-worth. On this account, envy is distinct from other emotions involving pain or distress (like jealousy or longing) in that the good in question is not the direct issue, but is rather, an occasion that prompts the issue of comparative self-worth. (This account of envy makes sense of the idea that the envious party is just as satisfied by the rival's loss of the good in question as it is with the subject's gain of an equal or greater comparable good.)

It should be obvious why this account of envy will not work for Velleman. His account of survivor guilt requires that envy be modified into envious-resentment through recognition of the survivor's *undeserved* good fortune. And it is precisely this lack of desert that does the work of explaining why it is rational for the survivor to feel normatively vulnerable – unable to justify his good fortune to others – and thus anxious over the prospect of being an apt target of envious-resentment. But, if envy is understood as distress over *deserved* good fortune, then the envious-resentment of others is unwarranted, and the survivor would be irrational to feel guilt over that prospect.

resentment undermines the project of explaining non-moral guilt. One might be tempted to think that appeal to resentment in *any* form (envious or otherwise), introduces moral considerations into the discussion of non-moral guilt. Comparison of the resentful modification of anger with the resentful modification of envy might appear to invite this reading. If emotions are ever modified in the way Velleman suggests, then non-moral anger can be modified into angry-resentment as a reaction to the discovery of wrongdoing. Thus, the original non-moral anger becomes a specifically moral emotion. One might then think that envy is modified by resentment in the same way, as we saw in the Mafioso example above. The envier initially feels non-moral (or immoral) envy over a rival's good fortune, but when it is discovered that the good in question was *unjustly* obtained, a moral judgment enters the picture whereby envy is modified into envious-resentment. Thus, envious-resentment is a response to wrongdoing, and an explicitly moral emotion.⁵⁹

If this were the case, Velleman's project of defending the rationality of survivor guilt would never get off the ground. Velleman is attempting to identify an emotion that is appropriate for the victims of catastrophe to feel towards the survivor, such that the survivor's feeling of guilt is a warranted and rational response. If envious-resentment

⁵⁹ My intuition is that resentment *is* a specifically moral emotion that targets the quality of a person's will and is a response to the lack of regard or standing that we are shown by that person through a specific action. If this is the case – if resentment is always a response to the quality of will manifest in an action – then Velleman's notion of envious-resentment is an inappropriate response to the survivors of misfortune. The survivors he is interested in did not act to procure their survival and so did not demonstrate any quality of will, good, bad, or otherwise. Thus, resentment lacks its proper target and is out of place. I grant this point to Velleman since I don't believe it is necessary to settle the question whether resentment can be nonmoral (i.e. directed at something other than the quality of a person's will) to show that Velleman's account fails to defend the rationality of survivor guilt.

turns out to be a moral emotion, then the proper target of envious-resentment is a moral wrongdoer, and the appropriate response by the wrongdoer is the feeling of *moral* guilt. But in cases of survivor guilt it would be a mistake on the part of the victims of a catastrophe to judge that the survivor is morally culpable, and this would make it irrational for the survivor to feel guilt as a response. Thus, if envious-resentment is always a moral emotion, and is indeed what the victims of catastrophe feel towards the survivor, then Velleman not only fails to show that non-moral survivor guilt can sometimes be rational, he in fact succeeds in showing that it is always irrational.

But Velleman would object that this criticism loses track of his general notion of normative vulnerability and the claim that the normative content of such vulnerability can be either moral or non-moral. (156) Velleman would point out that envious-resentment *can* have moral content, and thus *can* be a moral emotion, but it can also have non-moral content, and in these cases is *not* a moral emotion. Thus, in the case of survivor guilt, envious-resentment can be understood as a warranted non-moral response by others to the survivor's violation of some non-moral normative standard, such that it makes sense to speak of the target of this response as "unjustified" and "unjustifiable" in his possession of an "undeserved" good, with no standing to complain about this response.

To support the idea of non-moral normative vulnerability, Velleman must point to cases where the good at issue between the envier and envied is not illicitly gained (as in the example I gave above), but is nevertheless "undeserved" in some sense. He might point to cases like a chronic invalid's envy of a naturally healthy person, or an

unattractive person's envy of someone naturally beautiful. The beneficiaries of these goods did nothing immoral or ill-intentioned to gain these goods – they were simply born that way – yet there is no sense in which they can be said to deserve those goods, which makes them possessors of undeserved goods.

6.4 Critique of Velleman: Desert, Justification & Normative Standards

Velleman's account of non-moral normative vulnerability turns on the claim that (non-moral) envious-resentment can be a warranted response, when there is the recognition that someone possesses an undeserved good which puts them in violation of a (non-moral) normative standard. His defense of the rationality of survivor guilt depends on this being precisely the dynamic that takes place there. My next step is to determine whether survivor guilt actually fits this account. To do so I will compare survivor guilt to another case involving non-moral envious-resentment: The case of the chronic invalid's envious-resentment of the naturally healthy person. The question is whether the naturally healthy person can be understood as possessing an undeserved good that puts him in violation of some normative standard in a way that warrants the chronic invalid's envious-resentment. Through examining this case I will try to show that various possible uses of the notion of "desert", including at least one important non-moral use, fail to capture the dynamic by which the naturally healthy person is supposed to feel normative vulnerability over the mere possession of good health. I will then argue that this shows what we might have suspected to begin with: There is no clear, plausible non-moral normative standard for possessing good health that could be violated, and so no plausible

sense in which the naturally healthy person could be judged to have an undeserved good. Lacking the claim that the naturally healthy person possesses an undeserved good makes the envious-resentment of the chronic invalid unwarranted, and thus the healthy person's anxious guilty feeling cannot be explained as rational. I will then argue that, since this case of non-moral "guilt" provides a strict parallel with survivor guilt, it shows that survivor guilt fails to meet the criteria of Velleman's general account of guilt, and so cannot be defended as rational on those terms. Finally, I will suggest that the source of Velleman's error is the conflation of two senses in which one may "lack justification" for a good, both of which leave the possessor with no way to defend possession of that good, but only one of which entails a failing on the part of the possessor. Failure to distinguish these senses leads Velleman to see goods like survival as undeserved, where they are not.

6.4.1 The Chronic Invalid's Envious-Resentment of the Healthy Person

Imagine two people: one naturally healthy, the other a chronic invalid. The naturally healthy person is never sick and never has to do anything to maintain his health, while the invalid is routinely ill and must keep to a strict regimen to avoid aggravating his condition. To keep moral considerations out of the picture, also imagine that the naturally healthy person does not enjoy any other advantage over the invalid in terms of money, time, better healthcare providers, or other resources that contribute positively to his state of health. He's just naturally healthy. The upshot is that the naturally healthy person possesses a good that makes his life an overall more pleasant and enjoyable one – a good that the invalid lacks, and one that we can imagine the invalid envying. Since the

naturally healthy person has done nothing at all to gain his particularly robust physiology, it could be said that he does not deserve his good fortune. On Velleman's account, it seems that if the invalid recognizes the healthy person's good fortune as undeserved, his initial envy will be modified into envious-resentment. The naturally healthy person then finds himself in the position of being unable to justify his good fortune, and in recognizing this normative vulnerability, is rational to feel the anxiety of guilt as a response to the warranted envious-resentment of the invalid.⁶⁰

The invalid's envious-resentment of the naturally healthy person seems to provide a parallel to the phenomenon of survivor guilt. Both the survivor and the healthy person possess goods that the victim or invalid lack. Neither the survivor nor the healthy person have done anything to gain the goods in question, so neither can be said to deserve those goods. (And since no actions are involved, this is obviously not a moral issue.) Recognition that the survivor or healthy person possess undeserved goods is said to warrant others in feeling envious-resentment over their good fortune. And recognition that they are unable to provide a justification for these undeserved goods to others makes it rational for the survivor or healthy person to feel the anxiety of guilt as a response to warranted envious-resentment.

⁶⁰ To keep this case completely free of moral content we also have to imagine that the invalid's envious-resentment of the naturally healthy person is over the *sheer fact* that the healthy person enjoys robust health, and is not a response to the healthy person's failure to appreciate or utilize that health in appropriate ways. If the invalid's envious-resentment is over the fact that the healthy person does not make proper use of their good health (sitting around playing video games all day), or is over the fact that the healthy person does not seem to fully appreciate the worth of his health, then it seems that we are introducing moral content into the picture. If it is not exactly immoral, then it at least seems un-virtuous to misuse or fail to appreciate a good in this way, and thus one *might* be seen as an appropriate target of resentment.

With this parallel to the case of survivor guilt in place, we are in a position to examine the claims that are central to both, and in the process show why normative vulnerability and guilt are not justified in either.

Velleman's account would have it that the naturally healthy person, like the survivor, enjoys a good that he does not deserve, which leaves him unable to justify himself to others, and makes it rational for him to feel guilt over this fact. The key claim in both cases is that the good in question is *undeserved*. It is only recognition of an undeserved good that warrants the modification of envy into envious-resentment on the part of the invalid or victims of disaster. And, it is only the healthy person's or survivor's recognition that they possess undeserved goods that creates the feeling of normative vulnerability that makes guilt a rational response.

The problem is that it is not clear what it means in these cases to say that health, or survival, is an undeserved good that cannot be justified to others. Both health and survival seem to be things that are *per se* good to have, and *per se* bad to lack, but not the kinds of things that one could deserve or not deserve (at least not without introducing moral considerations). Velleman clearly intends for "undeserved" to be understood as having normative, but non-moral, content. But when we consider various senses in which a good could be understood as "undeserved" we see that these senses either introduce factors not present in the kind of survivor guilt Velleman wishes to defend as rational, or they introduce a notion of "undeserved" to which demands for justification do not apply. And, if the survivor cannot be understood as unjustified in his possession of a

good, he cannot be in violation of a normative standard, and any guilt he feels cannot be defended as an instance of normative vulnerability.

Let's consider what it could mean to say that a good like health (or survival) is something that is undeserved and unjustifiable.

6.5 Undeserved Goods: Having a Good One Should Not Have

We can begin our attempt to understand how a good could be undeserved and unjustifiable to others by thinking of that good as “something that one should not have”. But there are several different senses in which we could understand this locution.

One way for a good to be “something that one should not have” is for that good to have been obtained through illicit or unfair means. If a good has been obtained illicitly (through, say, theft, coercion, deception, etc.) then one will lack justification to others for possessing that good. Here we can make sense out of the claim that the thief, cheat, con-artist, or mugger does not deserve to have the goods obtained through those means. But this can't be the invalid's complaint against the naturally healthy person. The naturally healthy person has done nothing to obtain the good of health, so cannot be accused of using illicit means to do so. Moreover, if this was the case, then the issue would be a straightforwardly moral one and would not provide a suitable parallel with non-moral survivor guilt.

But it seems that we do use the term “undeserved” to pick out other ways in which a good could be “something that someone should not have” that do not turn on the good in question being obtained through illicit means. Imagine a sixteen year-old who receives a

Ferrari for her birthday. Let it be the case that neither the birthday girl nor her parents did anything illicit to obtain either the car or the money to purchase the car. We can imagine someone, well aware that the car was not obtained illicitly, still claiming that no sixteen year old deserves a Ferrari. If not deserving a good is taken to mean that one has a good that one should not have, then how could a Ferrari be a good that someone should not have?

There are ways that the Ferrari could be a good that the sixteen year-old should not have that can be ruled out because they obviously do not involve issues of desert. Someone might believe that a Ferrari is something that a sixteen year-old should not have, because Ferraris are incredibly fast vehicles, and sixteen year-olds are typically immature, inexperienced and irresponsible, and the combination of the two is a recipe for one dead or seriously injured teenager (or innocent bystanders). But this is a claim that the Ferrari is *too dangerous* for a sixteen year-old and that she shouldn't have one *for her own good* (or for the good of others) – not that she shouldn't have one because she doesn't deserve it. Similarly, the judgment that the sixteen year-old has a good that she should not have could be taken to mean that giving a sixteen year-old a Ferrari amounts to “spoiling” them in some way. Receiving extravagant gifts at a young age doesn't prepare a person for the reality of having to work for a living and pay one's own way – it makes things too easy for them, sets them up for failure later in life, etc. This too would

be the claim that the Ferrari is bad for the sixteen year-old, not that the Ferrari was undeserved.⁶¹

To rule these possibilities out, imagine a scenario in which neither the safety nor the character development of the sixteen year-old is the issue. The sixteen year-old in question is Dale Earnhardt Jr.'s daughter (Dale III) who has been driving incredibly fast automobiles since grade school. She's incredibly well-trained as a driver, and moreover, a very responsible person, so no danger to herself or others. Moreover, she'll never have to worry about money. In addition to her considerable trust fund (she could live luxuriously on the interest for the rest of her life, should she need to), and what she stands to inherit, Dale III has already displayed an aptitude for racing and a keen business sense, such that she already has various promising career options ahead of her, so she won't be "spoiled" by the gift. Having ruled these considerations out we can still imagine someone making the judgment that Dale III does not deserve a Ferrari – that the Ferrari is a good that she should not have. The question is: a) what meaning does "deserve" have in this context and b) does this meaning capture what is claimed by the invalid about the naturally healthy person in a way that is supportive of Velleman's account of survivor guilt?

There are several ways of interpreting "deserve" that obviously will not work for Velleman. The claim that a sixteen year-old does not deserve a Ferrari cannot be an

⁶¹ Alternately, I suppose, either of these cases could be understood as involving the claim that the sixteen year-old does not deserve to be treated with such careless disregard for their physical or psychological well-being by the person bestowing such a dangerous gift. But then the claim is that the sixteen year-old doesn't deserve to be treated that way, not that the sixteen year-old doesn't deserve a Ferrari.

objection to the mere fact that someone received a good that they did not *themselves* pay for – that would be an objection to gifts *per se*. The person who says that Dale III does not deserve a Ferrari is unlikely to also object to her receiving a cake that she didn't purchase herself. Applied to the case of the invalid and healthy person this would amount to the claim that no one deserves to have health that was the result of something they themselves did not do. This would amount to the claim that the healthy person did not “pay” for his health in some way, so he doesn't deserve it. If that were the case, it would seem that the invalid should have the same complaint if his own condition spontaneously, and naturally, improved, or if he was given treatment, or a cure for his condition, that he himself did not develop or administer. It seems unlikely that Velleman is making this claim.

Nor can this be an objection to the mere fact that someone received a good that they themselves could not afford. Most of the children of the world receive goods on a regular basis that they themselves could not afford. It seems implausible to think that the person who says that Dale III does not deserve a Ferrari, also thinks that children do not deserve food, shelter, medical treatment, etc. simply because they themselves lack the money to pay for such goods. Applied to the the case of the invalid and healthy person this would amount to the claim that the healthy person possesses a good that he would not be able to obtain for himself if he lacked that good. The healthy person could not afford to gain good health if he lacked it, so doesn't deserve it. Again, by analogy, this seems to entail that the invalid ought to object to any treatments or cures that he himself could not afford,

which (depending on his financial situation) might rule out many treatments covered by his health insurance. Once more, it seems implausible that Velleman would endorse this claim.

More plausibly, the claim that Dale III does not deserve a Ferrari could be understood as an objection to the fact that some people possess extravagant wealth and goods, while others work just as long and hard, yet can only afford much less. There is a discrepancy between the work-to-benefit ratio for people like the Earnhardts and the rest of us, that simply isn't fair. No one deserves to revel in such luxury while others toil to gain the necessities of life.

But this is a complaint against an unjust socio-economic system that is both moral in nature (and thus unsuitable to help explain non-moral guilt) and does not fit the case of the invalid and healthy person. In the case of wealth we have a shareable resource that (arguably) some are able to possess at the expense of others. If there is only so much money to be had in a particular economy, then one person's possession of a large chunk of that good means that others have less of it. But health itself is not a shareable resource that one person could hog-up at the expense of others (and if it was, or if we were discussing hoarding the *means* to good health, we would once more be dealing with a moral issue and no closer to explaining the rationality of non-moral guilt). Thus, the naturally healthy person's possession of the good of health cannot be understood as at the

expense of the chronic invalid's possession of that same good.⁶² If the invalid is making this claim, then it is both a moral one, and based on a gross misunderstanding of the nature of health, that would make the health person's feeling of guilt completely misguided.

We can, however, imagine a case where the claim "the Ferrari is undeserved" is not motivated by a complaint against an unjust socio-economic system. We might imagine that the person making this claim accepts the existing economic system as completely fair, and has no *general* objection to the fact that some people end up with fabulous wealth and Ferraris, while others do not. Here the claim that the Ferrari is undeserved could be based on the particular features of its recipient, such that the good of a Ferrari is seen as inappropriate, because it is not properly appreciated or utilized.

Imagine that Dale Jr. has twin daughters, Dale III and Dale IV. He gives both girls Ferraris on their sixteenth birthdays. But despite her family heritage, and despite having received the same early-childhood drivers education as her sister (she drives as well as any average driver, and responsibly, so is not a danger to herself or others), Dale IV simply does not share the same passion for high-performance automobiles. So, while Dale III is out at the track in her spare time, enjoying everything that a Ferrari can do, Dale IV is using her Ferrari to drive to the library, the local museums, and the local

⁶² Nor will it do any good here to switch targets and say that invalid is really claiming that *he* does not deserve to have *poor* health – everyone deserves to have good health, so he lacks something that he in fact deserves. In this case, if everyone deserves to have good health, then the invalid would have to admit that the healthy person has something that he deserves. Thus, the invalid can only be envying the naturally healthy person over deserved good fortune, and cannot be understood as feeling envious-resentment. As a result, the naturally healthy person has no reason to feel guilt (normative vulnerability).

college where she likes to study. Her Ferrari is rarely in anything over third gear, averages a daily speed of about forty m.p.h., and spends most of its time in one parking lot or another. Moreover, the difference between Dale III and Dale IV is not just one of preferences. Despite her training, etc., Dale IV cannot really feel the difference in handling and performance between a Ferrari and a Honda Civic and just doesn't "get" why others are so worked up about Ferraris. Thus, she also lacks the sensibilities that would enable her to handle a Ferrari in a way that would exhibit its exceptional qualities.

Now imagine that Dale III and Dale IV meet a car aficionado (let's call him Earl). Earl learns that the sisters own Ferraris and develops a serious case of envy toward both. As he gets to know the girls he realizes that Dale III understands and appreciates everything about her car and knows how to put it through its paces. Earl concedes that, given her skills and discernment, she deserves to own a Ferrari (or, he at least does not judge that she doesn't deserve to own a Ferrari). But Earl also realizes that Dale IV cannot discern the difference between the performance capabilities of a Ferrari and lesser vehicles, and that she does not appreciate the sheer awesomeness of a Ferrari. He comes to the conclusion that she doesn't deserve one. Thus, he feels envy over Dale III's possession of a Ferrari, but feels envious-resentment towards Dale IV.

Here Earl's complaint that Dale IV does not deserve a Ferrari is not a moral complaint. Earl does not think that either Dale IV or her father procured this good through the use of illicit means, or via an unjust economic system. Nor does Earl believe that Dale IV is danger to herself or others through her possession of this good. He does

not object to Dale IV (or anyone else for that matter) receiving gifts in general, nor to receiving gifts outside of one's own financial means. The reason he has for claiming that Dale IV does not deserve the good of a Ferrari seems to be a kind of aesthetic judgment: that individuals who do not (and possibly cannot) fully appreciate and utilize certain goods really shouldn't have those goods – they should be in the possession of someone who can and will fully appreciate them. (We can imagine Earl thinking: "It's a damn shame that a fantastic car like a Ferrari is being wasted on someone who just doesn't get how fantastic it is.")

It seems plausible that the world of car aficionados share some standard (however debated and debatable) that determines what counts as automotive excellence, driving ability, and an individual's ability to discern those qualities themselves. Thus, cars and individuals can be judged as either measuring up, or failing to meet that standard. On the basis of this standard, we can imagine a car aficionado judging that Dale IV has an excellent car, but that Dale IV lacks the ability to recognize or appreciate the qualities that make that car excellent, and the sensibilities that would allow her to bring those qualities out in the car, and because of this, Dale IV does not deserve to have a Ferrari.

In turn, we can imagine Dale IV feeling normative vulnerability in this situation. As she begins to recognize how highly Earl and other car buffs judge the quality of her vehicle, and the (non-monetary) value they place on it, as well as how poorly her own driving skills measure up to it (how little she is able to appreciate its qualities compared to others who could, but who lack the means) she too could come to believe that she

doesn't really deserve such a good – that it is, in fact, wasted on her. As she comes to realize that she can give no justification for her possession of that good, she might consider herself to be in violation of the normative standard held by Earl and other car buffs, and come to feel anxiety over the prospect of their warranted envious-resentment.⁶³

The case of Dale IV (and its sense of undeserved) comes closest to capturing what Velleman needs for his account to work. The judgment that a good is undeserved appeals to a normative standard that can be free of moral content. It also seems possible for both the envier-resenter and envied-resented to recognize and share this normative standard, and thus to make sense of how someone could be judged, and come to see themselves as, in violation of that standard. Thus it seems possible for someone to feel unable to justify the possession of some good on non-moral grounds and to feel non-moral guilt over the warranted envious-resentment of others.

But, however much the Ferrari case provides us with a plausible example of non-moral “guilt” (I’m not convinced that something like *embarrassment* isn’t the better word, but I won’t argue that here), and a non-moral use of the word “undeserved”, this case still fails to fit the case of the naturally healthy person and chronic invalid in a

⁶³ If the Ferrari example does not resonate, imagine someone who owns several original paintings by Monet. The paintings were purchased for a fair price from Monet himself, have been passed down by the current owner’s family for generations, and now hang in her living room. She meets an art historian and collector who envies her possession of these paintings. The collector then learns that the current owner suffers from blue-green color blindness and cannot distinguish the subtleties of Monet’s palette. The collector has no objection, in principle, to the private ownership of great works of art, nor any objection to inherited wealth or goods. He only believes that such works should at least be owned by those who can actually discern and appreciate their qualities. On that basis he comes to the conclusion that the current owner does not deserve to own these paintings and feels envious-resentment toward her. It is easy to imagine the owner of the paintings recognizing that she lacks the capacities that would allow her to fully appreciate these goods and, seeing herself as the apt target of the collector’s envious-resentment, feeling normatively vulnerable and anxious.

significant way. When Earl judges Dale IV to be undeserving of a Ferrari he is responding to the fact that Dale IV is unable to fully appreciate, or make use of, the good she possesses. If this was the nature of the invalid's issue with the naturally healthy person, the complaint would be that the healthy person fails, or is unable, to appreciate or make appropriate use of his good health. And, if this is the case, the invalid's judgment is based on what the healthy person does (or fails to do) with his good health, and not on the sheer fact that he possesses that good *per se*. As a result, the invalid's envious-resentment of the healthy person does not provide a precise parallel with survivor guilt.⁶⁴ The version of survivor guilt that interests Velleman is one where no actions, or inactions, of the survivor are at issue: the survivor feels guilt over the *sheer fact* of having survived misfortune where others did not – not over what he does, or fails to do, with his life after the misfortune. So while the Dale IV case might show that something like non-moral guilt is a rational response to certain forms of non-moral normative vulnerability, it does not show that the form of survivor guilt Velleman is interested in – where the actions or inactions of the person are not an issue – can be defended as rational.

6.6 Survivor Guilt Revised?

But what if the Dale IV case was taken to provide a more accurate analogy with the phenomenon of survivor guilt? Could Velleman accept the claim that survivor guilt is better understood as anxiety over what one does, or fails to do, with the good of survival, rather than anxiety over the sheer fact of survival, and then go on to argue that the Dale

⁶⁴ Recall also that this kind of situation was ruled out earlier to keep issues of morality, or judgments about virtue from creeping into our scenario.

IV case shows how such anxiety can be non-moral and rational? On this revised account of survivor guilt, the victims of disaster judge that the survivor is not meeting the normative standard for “appreciating and making appropriate use of one’s survival” (let’s call this “living life well” for short) and since there are others who would have made better use of that good, but were denied the chance, the survivor does not deserve the good of survival. The survivor comes to accept this judgment and, with nothing to say in his own defense – nothing that would justify his possession of that good to others – feels the anxiety of normative vulnerability in response to the warranted envious-resentment of others.

Defending “survivor guilt revised” would be a difficult task that would necessarily involve one in making controversial claims about non-moral standards for “living life well”, as well as claims about how that standard could be violated in non-culpable ways that make guilt rational, because they leave one a target of warranted envious-resentment. I want to suggest that attempting such a task will fail. I think that Velleman’s original understanding of the phenomenon of non-moral survivor guilt, where the actions (or inactions) of the survivor are not at issue, is the correct one, and any attempt to defend something like the revised account articulated above picks out a phenomenon that should not be understood as survivor guilt.

To show this I will grant that there is some non-moral normative standard by which we are able to judge whether someone is “living life well”, that it is possible to be in non-culpable violation of that standard, and that feeling guilt over this failure can be rational.

Imagine now that both the victim and survivor recognize this normative standard. The victim claims that the survivor is in violation of that standard, because he fails to make proper use of, and have an appropriate sense of appreciation towards, the good of his own life, where, by contrast, other non-survivors would have made better use of that good. The survivor accepts this judgment, and with nothing to say in his own defense, feels guilty for having an undeserved good. My intuition is that this phenomenon is not survivor guilt.

Being concerned and anxious over how to live one's life, is a universal feature of human existence. We each raise this issue for ourselves, in varying degrees of explicitness and perspicuity, and in varying degrees of frequency and intensity, when we confront our own finitude and mortality, and recognize that we alone bear responsibility for what we make of our lives. In some fashion we each raise for ourselves the question of what to become, and we answer that question through our choices and actions. The anxiety we feel over how we are living our lives and what we are making of our selves in the face of our own finitude and mortality seems best described as "existential anxiety" (or "existential guilt").

But, being anxious over how one's life is to be lived, and feeling guilt over the suspicion that one is not living that life well, is hardly unique to the survivor of misfortune. This is an anxiety and guilt that we can all experience and share in. What *is* unique to the survivor is an experience that has harshly forced him to consciously and explicitly confront the same questions that we all face. The fact of survival is then

merely the occasion that forces the survivor into a reckoning with his own finitude and mortality, which reveals the anxiety that is at the root of human existence itself. Thus, the proposal that the survivor feels guilt over failure to meet a normative standard for “living life well” – however non-moral, non-culpable, and rational we may grant this to be – does not pick out a phenomenon that is unique to surviving a disaster, and so cannot properly be called survivor guilt.

Velleman’s general account of guilt, as normative vulnerability, works well to explain the phenomenon of moral guilt, and, I believe, can be extended to cover some forms of non-moral guilt. But guilt of either form can only be normative vulnerability if there is a violation of some normative standard. In our examination of the chronic invalid’s (alleged) envious-resentment of the naturally healthy person, we saw that none of the examined non-moral uses of “undeserved” preserved the dynamic that takes place in Velleman’s account of survivor guilt. This shows that there is no sense in which the healthy person should come to see the good of health itself as undeserved, which leaves him with no reason to see himself as the apt target of warranted envious-resentment. I take the inability to make sense of how health could be seen as undeserved to show that there is no plausible non-moral standard for having naturally good health that can be violated, and so no reason for the healthy person to feel unjustified, and hence no reason for the healthy person to feel normatively vulnerable and guilty.

If the case of the chronic invalid and healthy person provides a good parallel with survivor guilt (and I think that it does), then this shows that there is no good sense in

which the survivor could take survival itself to be undeserved, and so no reason for the survivor to see himself as the apt target of warranted envious-resentment. Again, the obvious reason for this is that there is no plausible non-moral standard for surviving catastrophes that can be violated, and so no reason for the survivor to feel unjustified, hence no reason for the survivor to feel normatively vulnerable and guilty.

6.7 Lacking Normative Standards & Lacking Justification: Where It All Went

Wrong

But isn't it the case that neither the healthy person nor the survivor can give justification for their possession of their respective goods? And isn't lacking justification for the possession of a good enough to show that that the good is undeserved, and that the possessor of that good should feel guilty?

The crux of the problem in these cases is that it confuses two senses in which one may "lack justification". There is the more familiar sense that shows up when we lack moral justification. I deliberately injure you to eliminate your competition for a job. When confronted, I have nothing to say in my defense – nothing that does not reveal the ill-will behind my action. As we have seen in the Dale IV case (or the Monet case, if you don't like Ferraris (see fn.11)), there can be a sense in which one may lack non-moral justification. Being in possession of a good that one does not sufficiently appreciate, or cannot make proper use of, may leave one with nothing to say to defend oneself against the envious-resentment of those who do appreciate, and could make proper use of that same good. It is also possible to have unjustifiable beliefs. I might claim to believe that

all of our successful politicians are extra-terrestrials, but when asked to defend that claim I can only point to their success. I also might have unjustifiable emotions, such as a deathly fear of lint, although I cannot provide any reason why anyone (myself included) should fear lint.

But one may also “lack justification” where a call for justification simply does not apply. I am, at this time, a certain height. There may be an explanation for this fact (some story involving genetics, nutrition, etc.), but there is no justification for it. I was born with a specific nationality. This fact can be explained by the location of my birth and the nationalities of my parents, but it makes no sense to ask for a justification of my birth nationality. (Note that this is different than asking me to give a justification for not changing my nationality, and is also different than asking me to give a justification for why I claim that some specific nationality applies to me.)⁶⁵ Some people happen to be naturally healthy, while others are chronically ill. The person who possesses naturally good health lacks justification for that health, but not because he lacks a justification he ought to have, but because justification does not apply here. There are certain facts – one’s height, birth nationality, or natural physiology, for example – for which we are not

⁶⁵ Nor does the fact that I may have been advantaged in various ways by these facts alter the situation. It might be the case that my height gives me certain social advantages. It might be the case that my parents were kinder and more loving than others, such that I enjoyed a happier childhood and a more well-adjusted adult life. It might be a fact that I was born in a more peaceful and prosperous country than others and enjoyed those benefits. And, it might be the case that these facts bring with them certain responsibilities. I might be morally obligated not to use my height to intimidate and dominate others, or to be more understanding of those whose childhoods were less than happy, or to not believe that I am superior to other people because of my nationality. And if I do (or fail to do) any of these things, I might owe justification to others for so doing. But this is not a justification for the original facts themselves, but rather, for what I make of those facts.

responsible and so cannot be answerable. Any call for justification of these facts is inappropriate.

This is the source of confusion in Velleman's account of survivor guilt. He takes the survivor's inability to justify his good fortune to be a failure in the first sense of "lacking justification", rather than the second sense. He takes the survivor to be under a requirement to provide justification for something that is non-justifiable, and sees his failure to do so as an inability to provide what he ought to be able to provide. But the survivor is under no such requirement. Surviving a catastrophe, where no relevant actions are involved, is something that happens to some people, and not to others. It is a fact that might have an explanation, but it cannot require justification. Mere facts are non-justifiable – to ask for their justification is to be confused about the very notion of justification. One might as well go on to ask the survivor to justify his current height, genetic background, and place and time of birth. The survivor did nothing to bring about these facts, or the fact of his survival, so there is no sense in which he could be called upon to give justification for any of them. So, while Velleman is correct to say that the survivor lacks justification for his survival, this is not a failure on his part. The survivor does not lack justification where he ought to have one – there is nothing he could say that would count as justification.

6.8 Concluding Observations on Velleman's View

David Velleman defends the rationality of non-moral survivor guilt by claiming that the dynamic found there fits his general account of guilt as normative vulnerability.

Normative vulnerability occurs when one finds oneself in violation of some normative standard and unable to justify oneself to others on terms they need accept, thus making the anxiety of guilt a rational response to finding oneself the apt target of various negative responses. Velleman claims that since the victims of a catastrophe may be warranted in feeling envious-resentment over the survivor's undeserved good fortune, and since the survivor recognizes that he cannot justify his good fortune to the victims, or defend himself against this warranted response, he is normatively vulnerable and rational to feel guilt over this prospect.

In this paper I argued that Velleman's defense of the rationality of survivor guilt fails to meet the criteria set forth by his general account of guilt. To show this I constructed a parallel case of non-moral guilt: the case of the chronic invalid's envious-resentment of the naturally healthy person. This case provides an example of a person (the naturally healthy person) who feels normative vulnerability over the fact that he cannot provide a justification for his possession of the good of health – a good not acquired through any action of his own – which leads him to feel anxiety over the chronic invalid's warranted envious-resentment of this undeserved good.

My intuition in this case (and in the case of survivor guilt) was that there is no plausible non-moral normative standard for possessing good health (or surviving catastrophes) that can be violated, so no way to understand the healthy person (or survivor) as normatively vulnerable and rational to feel the anxiety of guilt. To show this I examined various ways in which we might use the term “undeserved” (drawn from the

Ferrari cases) and argued that any of these possible uses failed, because they either imported moral considerations, and thus were unsuitable to explain the phenomenon of non-moral guilt, or, on the one plausible non-moral use examined, did not involve normative vulnerability over the mere possession of the good of health, but rather, over the appropriate use or appreciation of that good.

I took this examination to show that the use of the notion of desert is out of place in the case of the chronic invalid and naturally healthy person, which, in turn, shows that there is no non-moral normative standard for possessing good health that could be violated, and so no sense in which the naturally healthy person could rationally feel normative vulnerability. As a precise parallel to the phenomenon of survivor guilt, this case shows that the same shortcomings are to be found there: There is no non-moral normative standard for surviving catastrophes that can be violated, and so no sense in which the survivor could rationally feel normative vulnerability.

To the possible response that our understanding of the phenomenon of survivor guilt be revised to reflect the dynamic of non-moral guilt discovered in the Ferrari cases, which I rejected as inadequate to explain the case of the chronic invalid and naturally healthy person, i.e. that survivor guilt could be understood as guilt over failure to properly appreciate or make appropriate use of the good of survival – I defended Velleman’s original description of survivor guilt as guilt over the sheer fact of survival. I suggest that any attempt to revise survivor guilt in this way leaves us with a term that fails to pick out anything unique to the survivors of misfortune. Anxiety over whether

one is living one's life well is a feature of human existence that is common to us all, not just the survivors of misfortune, and is better described as "existential anxiety". Thus, survivor guilt, as a unique phenomenon, is best understood as guilt over the sheer fact of survival, but can only be understood as an irrational response on Velleman's terms.

To explain why Velleman is led to see survivor guilt as rational, I suggested that the source of error is a failure to distinguish between two ways in which a person may "lack justification" or "have nothing to say in one's defense", one of which leaves a person normatively vulnerable, one of which does not. One is normatively vulnerable where one is answerable for one's actions or inactions, attitudes, beliefs, or emotions, and cannot provide a justification to others that one ought to be able to provide. But one is not normatively vulnerable for facts about oneself (one's height, race, birth nationality, etc.), or for facts about that which merely befalls one (having naturally good health, surviving a catastrophe through no action of one's own). Mere facts can be explained, but they cannot be justified. Thus, the person who is unable to provide justification for such facts does not fail to provide something that ought to be provided, but fails to be able to respond to a call for justification that is utterly misplaced. Thus, on Velleman's account, feeling guilty over one's survival cannot be defended as rational, since it involves a mistake by both the victims and the survivors. The victims treat the survivor as if he were answerable in some way for his survival and ask him to give a justification where none is possible, and the survivor feels normative vulnerability over his inability to provide the impossible.

6.8 Normative Vulnerability vs. Relational Vulnerability: A Nietzschean Counter-Proposal

I argued that Velleman's defense of survivor guilt fails to meet the criteria of his own general account of guilt, because there is no non-moral normative standard for surviving catastrophes, and so the claim that the survivor has an undeserved good is mistaken. I also argued that the source of error that is internal to his account is his failure to distinguish between two ways of "lacking justification," such that he conflates having nothing to say in one's defense, when one is answerable for one's actions, attitudes, beliefs, etc., with having nothing to say in one's defense because any request for justification is inapt.

However, I also think that Velleman's more basic mistake – a common one – goes deeper than internal inconsistency and is one of the guiding assumptions of his argument. This is the assumption that moral guilt is the basis upon which we should model all other so-called instances of "guilt." Morality has a particular structure in the context of which the concepts it employs make sense in relation to each other. In the context of morality we can make sense of the claim that a normative standard has been violated and another person has been wronged, and we can make sense of the claim that the anxiety of guilt is felt over wrongdoing and possible negative responses.

But imposing the structure of morality onto other forms of anxiety only serves to distort them into pseudo-moral phenomena. We saw this in Velleman's ineffective attempt to provide a normative standard that justifies survivor guilt. My worry is that, because

morality is so familiar to us, and has shaped our thinking, we have a tendency to think in moral terms without reflection and to apply those terms in unhealthy ways. I worry that this in fact exacerbates forms of anxiety that are already problematic, by misconstruing their nature through the use of moral terminology.

For instance, while Velleman may not intend this outcome, his use of moral terminology to explain survivor guilt, such as the claim that others are justified in their resentment of the survivor, cannot help but foster and encourage the survivor's feeling that he has in some way involved himself in something wrong. But with no normative standard to define the nature of that wrong, or to give some indication of how, and to what extent, the negative responses of others are justified, the survivor is left with a pervasive sense of anxiety. Even worse, since the "wrongness" in this case is ultimately over being alive, this introduces doubt into the survivor's sense of his own worth. If others are justified in resenting the survivor's "underserved" continued existence, then surely this implies to the survivor that he ought not exist.

This development is notable, since Velleman's account of moral guilt as "normative vulnerability over justified negative responses" does not appear to suggest that a sense of worthlessness necessarily accompanies the feeling of guilt. The wrongdoer can recognize that others are justified in mistrusting his motivations, but this need not involve the feeling that he is incapable of changing his motivations, or living up to moral standards,

and so need not imply that he is worthless as a moral agent.⁶⁶ Insofar as this is so, understanding moral guilt as “normative vulnerability” provides a way to understand moral guilt as a healthy, and appropriate psychological response to recognition of one’s own wrongdoing, that can motivate efforts towards moral development, apology, repair, and restoration of relationships.

But the attempt to impose concepts taken from the context of moral guilt onto the phenomenon of “survivor guilt” seems to reintroduce a pervasive feeling of anxiety and sense of worthlessness. I claim that this is because Velleman’s attempt to use the structure of moral guilt to explain “survivor guilt” is an instance of conflating ethical obligations with living up to ideal standards.

We saw the same kind of conflation occur in the creditor-debtor relationship to the ancestors in Chapter One. In that context, ritual obligations to the gods were turned into ethical obligations, because ritual failures were perceived as a threat to the entire community. Failing to repay the gods adequately through sacrifice, invites the anger of the gods, and brings the threat of harm to one’s neighbors, so religious failures become a matter of community concern. This means that failure to demonstrate proper gratitude to the gods becomes an ethical failure. Thus, the failure to live up to an ideal of virtue

⁶⁶Perhaps others feel differently and take the wrongdoer’s action to be evidence that he is to be mistrusted *tout court*. This is where the idea of justification matters. The wrongdoer may recognize that his action was wrong and that others are justified in mistrusting him to a certain extent, or with regard to certain matters, but should also be able to recognize when others are being unreasonable and are unjustified in their responses. In this case, the wrongdoer may recognize that he faces a hostile community that won’t trust him on any matter, and that his future prospects are bleak, but this is not an instance of normative vulnerability and so what the wrongdoer feels here might be called distress or apprehension, but it is not guilt or a sense of worthlessness.

(gratitude) is taken to have the same significance as wronging another person. Once this connection is firmly established in human psychology, and reinterpreted by the ascetic priests, failure to live up to ideals produces feelings of guilt and the idea that punishment is justified. Worse yet, it seems that through the influence of this interpretation on human psychology, the order of the relationship gets turned around such that any experience of suffering is interpreted as guilt, from which it follows that punishment (or the negative reactions of others) is justified.⁶⁷

I take this to be precisely what occurs in survivor guilt. The pervasive influence of morality on our thinking leads us to interpret anxiety and suffering as the feeling of guilt. When the survivor naturally feels anxiety over his survival (here it may be existential anxiety over how he has lived his life, prompted by his brush with death) he interprets this feeling as guilt, and from this interpretation the connection between guilt and punishment directs his attention to the negative responses of others, which must be justified, because, after all, he feels guilty. But this moral interpretation misconstrues and distorts another more basic phenomenon.

We can turn to Nietzsche to recognize this more basic phenomenon. If Nietzsche's account of the development of moral guilt is correct, then it is a late development in

⁶⁷ In conversation on this topic, Herb Morris suggested that the feeling of survivor guilt is something like an "existential defense mechanism." (My term.) It is a natural human tendency to resist the idea that dramatically life-altering events can occur against one's will and for absolutely no reason or purpose. The survivors of these events feel them as threats to their agency that undermine any sense of control they might have over their own lives. Feeling guilty is a way of taking control over those events. To feel guilt is to feel responsibility, therefore, one's agency must have somehow been involved. This is an intriguing psychological analysis of survivor guilt. If it is correct, then it provides support for my claim that moral guilt has significantly influenced, and distorted, the way that we perceive other forms of anxiety.

human history, and is itself an interpretation of already existing phenomena, which were originally non-moral. This is, of course, a basic premise of Nietzsche's genealogy. Guilt is an interpretation of non-moral debt consciousness. In particular it is an interpretation of the creditor-debtor relationship to the gods, which developed to exhibit features characteristic of guilt, including a pervasive feeling of anxiety over punishment, and a consuming sense of personal inadequacy to meet one's obligations. And while this relationship emphasized obedience to customs, this was motivated out of fear of the gods (perhaps also a desire for the community's continued success). Pleasing the gods was a purely economic arrangement, however much this might have been obscured through religious trappings, and so was clearly also a non-moral in nature.

But we should not forget that the creditor-debtor relationship was itself an interpretation of even earlier pre-existing relationships and practices. The creditor-debtor relationship interpreted more basic forms of trade and exchanges of goods between individuals; the relationship of the individual to the community; obedience to customs and the practice of banishment; and respect for the ancestors in contractual and economic terms. Thus, before there were creditors, debtors, contracts, and punishment, there were more primitive human relationships that could go well or poorly, and human responses to those eventualities.

We should pay particular attention to the way that early human relationships can go poorly, and the responses that followed.

We saw in Chapter One that natural selection favored human beings with stronger herd instincts, because it led them to remain in communities. We also saw that a community that was better able to coordinate the activities of its members, to allow them to cooperate in mutually beneficial ways, provided greater advantages to its members and had greater chances of survival. Cooperative endeavors required that customary behavior was established, followed, and enforced. This was accomplished, perhaps primarily, through the alpha figures, but was more effective when customs were enforced by the entire community. This required the members of the community to develop an instinct to obey the customs of the community. Obedience to customs was enforced through banishing recalcitrant and uncooperative members. I argued that this practice selected for those members that had a stronger instinct to obedience, as well as stronger instincts for reciprocity towards other members of the community, resulting in human beings who were highly attuned to the customs of the community, the expectations of the other members, and the consequences that follow when customs are broken and expectations are not met.

We have seen how the practice of punishment in the creditor-debtor relationship was an interpretation and justification of the practice of banishment, which in turn, was traced back to the community's treatment of its defeated enemies. Thus, the common feature that runs through the development of punishment is angry retaliation and the pleasure of doing harm to the cause of one's injury.

I argued that the creditor-debtor relationship led to the practice of punishment as economic relationships became increasingly important for the survival of the community, and it was recognized that creditors were, through “wealth,” a new source of power in the community, who must be appeased if the stability of the community was to be maintained. I also argued that the creditor-debtor relationship simply assumed the fact of human anger over injury and the desire to retaliate as a given that must be accommodated. This is reflected again in the community’s relationship to the ancestors, who are understood as potentially angered creditors who may respond with violence.

I want to suggest that what we see as the most primitive phenomenon, which is interpreted by the creditor-debtor relationship as debt consciousness, and in turn interpreted through the ascetic ideal as guilt consciousness, is relational dependency and vulnerability.

Human beings have always lived in communities and have depended upon their relationship with others for survival. This relationship of dependency leaves them vulnerable to any negative responses that may affect those prospects. The process that selected for instincts of reciprocity produced communities of individuals who are most sensitive to this vulnerability. Moreover, through the herd instinct, human beings are also keenly aware that their differences and non-conformity may set them apart from others and thereby raise suspicion and the likelihood of negative reactions. The result is human beings who are attuned to negative responses and inclined towards behavior that is meant to mitigate those responses. This is more basic than normative vulnerability.

The most primitive form of human community, according to Nietzsche, was the family unit. Here, the expectations that mattered were the demands of the alpha figure, who protected the family, but dominated the other members and commanded obedience through aggression and the threat of violence. The other members depended on the alpha's protection, so obedience was necessary for survival. So, the most basic human relationships involved obedience out of fear and vulnerability, and this was prior to identifying any pattern of expectations that might count as a primitive normative standard. The way that most members learned the "rules" of the family was through being raised in it. But as children, they felt vulnerability and dependence on their parents long before they had any idea that their parents had rules or expectations. Thus, the feeling of relational vulnerability occurs prior to any sense of normative vulnerability.

It seems, however, that relational vulnerability is more than just temporally prior to normative vulnerability. Relational vulnerability is also logically prior to, and the ground of, normative vulnerability. If human beings were not vulnerable to being harmed in various ways (physically, emotionally, psychologically, etc.), and did not depend on each other for survival, then there would be no need for rules and standards that are meant to protect the individual's interests, or the community as a whole. Morality would not be a concern, and normative standards of morality would not exist. Thus, it is only through human relational vulnerability that moral standards, and normative vulnerability with regard to those standards, come into existence.

Understanding relational vulnerability as the ground of normative vulnerability allows us to explain the prevalence of normative standards. Human practices reflect various human interests, and where we have human interests we have the possibility of success or failure, harm or benefit, satisfaction or disappointment, differences and conflict, and the need to manage expectations and responses by developing normative standards. Hence we see “ethical” standards for the relationship between members of a community, and legal standards to regulate contractual, economic relationships early on in human communities. In the relationship to the ancestors this becomes slightly more obscure, due to the superstitions and religious trappings that have accrued around it to become religious practices. But the relationship towards the ancestors began with the community’s attitudes towards the treatment of their elder members, where it was thought that their previous sacrifices and contributions to the community were owed continued compensation – even after they had ceased to contribute as they had in the past. So, stripped of primitive superstitions about powerful spirits and later religious interpretations, what emerges is a kind of “virtue standard” concerning how much gratitude, honor, or reverence is owed to those who have provided benefits in the past and are responsible for one’s present successes. But in each of these cases, the ability to be sensitive to those normative standards presumes that the individual cares about the relationships they concern, and thus feels vulnerable to the dynamics of those relationships.

Understanding relational vulnerability as basic also allows us to explain forms of anxiety such as “survivor guilt” without forcing those phenomena into distorting moral categories. As human beings, we are vulnerable and attuned to the negative feelings and potential responses of other people. The survivor of a disaster recognizes that he is a source of pain to certain people because he reminds them of their loss and sorrow. The survivor is not the cause of their loss (the disaster is), but his presence, as a reminder, is the cause of current feelings of emotional distress.

Even if the survivor recognizes that he was not responsible for the disaster or any other moral wrongdoing in the process, and also recognizes that there is no normative standard for surviving disasters against which the envious-resentment of the victims would be justified, he still recognizes that human beings naturally have negative responses to suffering, that he is a cause of that suffering and vulnerable in that regard, and is anxious over the prospect of his own suffering as a result. This means that *there is* a sense in which the survivor’s anxiety is rational. It is rational to be anxious over negative treatment by others, even when that treatment is unjustified. But explaining this in terms of relational vulnerability allows us to avoid treating the survivor’s anxiety as a form of guilt, which threatens to convince the survivor that he deserves to suffer the victims’ negative responses.

Moreover, interpreting “survivor guilt” in terms of relational vulnerability allows us to explain another puzzle about the phenomenon. Even if we judge that the victims have no justification for feeling envious resentment over the survivor’s “undeserved good,”

and that the survivor should not feel either moral or non-moral guilt, we would at least find it disturbing, if not objectionable, if the survivor only felt joy over his own survival in the presence of those who feel loss.

Velleman, it seems, would have to say that the survivor is not responsive to the normative standard that he has violated, and is insensitive to the justified resentment of others, which is analogous to the wrongdoer who does not recognize moral standards and is insensitive to the moral standing and justified moral protest of those who were wronged. There is something correct in this analogy and the intuition that lies behind it, but it is obscured by the moral terminology and concepts used to provide the analogy.

We expect the survivor to feel some kind of distress or pain over this situation that goes beyond any self-referential existential crisis that might be prompted by the experience, and is, rather, responsive to the feelings of the victims. Explaining this in terms of relational vulnerability, we might say that we expect the survivor to be sensitive to the relational “rupture,” or “damage,” that has occurred between himself and the victims. We expect the survivor to be sensitive to the fact that he is a source of pain to others, and that this matters – that their negative responses are a natural reaction to loss. Here, I am not claiming that the survivor ought to endure retaliation from the victims or other extreme responses. I am only claiming that we expect a normal human being to be

responsive to the fact that the pain caused by his presence strains the nature of his relationship with the victims in a completely understandable way.⁶⁸

Thus, thinking in terms of relational, rather than normative vulnerability allows us to explain forms of anxiety like survivor guilt in a way that that does justice to the underlying feelings, without distorting them into pseudo-moral phenomena.

Understanding the survivor as anxious over relational vulnerability allows us to clarify what count as reasonable responses on the parts of the victims, and what seem like reasonable responses on the part of the survivor. Moreover, by identifying relational vulnerability as the underlying phenomenon upon which normative standards are constructed, it seems that we should be in a better position to evaluate and critique those standards in light of how they impact our relationships to each other, and may be an important first step in the self-critique and self-overcoming of morality.

68 Take another kind of case: Brad moves to a new neighborhood and meets a new neighbor, Sarah. He later discovers that he is the spitting image, and also remarkably similar in personality and temperament, to the now deceased John, who was the one and only love of Sarah's life. Every time she sees Brad, Sarah is reminded of John and sinks back into despair and grief. Over time, she becomes convinced that being with Brad would make her happy again, but Brad is not romantically attracted to Sarah, which causes her even more despair. In this case I assume that no one would judge that Brad is guilty for looking and seeming like John, or that the fact that he causes Sarah pain indicates that he owes her a chance at a relationship, or that he should move away so as not to cause her further pain (assume that the costs of doing so would be prohibitive for Brad). But it also seems that we would judge Brad to be insensitive and callous if he were not bothered by the fact that he causes Sarah so much pain, and that it would be objectionable if he did not try to avoid unnecessary contact with Sarah, knowing of the pain this causes.

Conclusion: The Elimination of Guilt, Punishment, and the Economics of Suffering

The main purpose of this dissertation was to defend the thesis that Nietzsche's rejection of the concepts of guilt and punishment is plausible, and that this rejection does not entail the objectionable consequences that would follow by leaving crime unaddressed or by eliminating any feeling of pain over wrongdoing. In order to defend this thesis it was necessary to closely examine Nietzsche's views on the development of debt consciousness, the bad conscience as consciousness of guilt, punishment, justice, and other views necessary to clarify his position on these topics. In the process of examining Nietzsche's views I made a number of arguments to support my own interpretation of the text. I will first review the arguments for the interpretation of Nietzsche that I have defended, and will then turn to the arguments that provide support for the claim that guilt and punishment can, and should, be eliminated from "psychology, history, nature, the social institutions and sanctions."

Chapter One presented Nietzsche's account of the development of debt consciousness through the influence of economic thinking in the context of the creditor-debtor relationship. In this presentation I defended interpretations of several challenging claims in the text of GM II.

The first interpretive issue was Nietzsche's claim that the morality of custom and creditor-debtor relationship were responsible for the development of the conscience, by creating a memory in human beings that permitted them to promise. The issue is that Nietzsche's description of the creditor-debtor relationship's means for making a memory,

mentions promising, obligations, and the conscience. Thus, it seems that the conscience, feelings of obligation, and making promises existed before the memory was created.

Since Nietzsche is proposing an account of the non-moral origins of moral phenomena, he seems to presume moral phenomena in his explanation, making it question-begging.

The solution was to point out two different senses of memory and promising, and two different factors in the development of the conscience and the feeling of obligation found in the text.

The text has two different senses of “memory.” There is the sense of memory that is the ability to recall past events, and an “active” sense of memory – the true memory of the will – that is the capacity for maintaining some feature of the past (such as a promise) as an ongoing source of motivation. Nietzsche’s language can give the impression that he offers an account of the memory as such, but this is not the case. Nietzsche’s only aims to provide an explanation of the “true memory of the will,” which allows a person to maintain the motivation to fulfill an obligation undertaken in the past.

The same is true of Nietzsche’s mention of promising. We can wonder about the origins of the human *convention* of uttering the words “I promise,” or we can wonder about the *ability* to commit oneself to a future course of action. Nietzsche does not offer an account of the development of the convention of promising, he offers an explanation for how promising became a source of obligation and motivation for human beings, through the use of punishment in the creditor-debtor relationship.

Nietzsche's mention of the "conscience" and "obligation" in the creditor-debtor relationship required recognizing that the two aphorisms typically taken to present the same events are in fact accounts of two different developments in the early human community. Nietzsche claims that the memory was created through the mnemonics of pain. GM II:3 and GM II:5 both present the use of pain to create a memory. GM II:3 can be read as offering general comments on the mnemonics of pain, and GM II:5 as illustrating the precise (and only) way pain was used to create a memory. But closer examination of the text shows that GM II:3 explains how pain was used to create a memory of basic rules for social co-existence in a community, and GM II:5 explains how pain was used to create a memory of a promise made and the obligation to fulfill it. I argued that GM II:3 is an account of the development of the "formal conscience" through the morality of custom, which impresses an obligation on the individual to follow the basic rules of the community. This is the most primitive sense of "conscience" in Nietzsche's account, and existed long before the development of the creditor-debtor relationship, as a necessary condition for the existence of any human community.

Nietzsche's mention of the conscience in his presentation of the creditor-debtor relationship is the formal conscience. GM II:5 describes the way that promising was impressed onto the formal conscience as a new source of obligation, made necessary in the context of the creditor-debtor relationship. Within this context individuals acquire a true memory of the will, which eventually becomes the fully developed conscience of the sovereign individual.

Nietzsche claims that the creditor-debtor relationship is the “oldest relationship between persons” and is “older than any societal associations or organizational forms.” On the surface, both of these claims are implausible.

First, human beings clearly related to each other in a number of ways prior to the development of contractual relationships, so this cannot be what Nietzsche means. I resolved this issue by pointing to the distinction between “human beings” and “persons.” Nietzsche does not claim that the creditor-debtor relationship is the oldest relationship between “human beings,” he claims that it is the oldest relationship between “persons.” I argued that Nietzsche uses the term “persons” as a normative category marking a distinct stage of development in the human animal. Personhood, in Nietzsche’s terms, is the ability to think in terms of values and reasons – to be capable of reasoning, reflection, abstraction, and imaginative thinking, and to guide one’s actions through these abilities. It is through the economic activities of gauging values, assessing equivalents, setting prices, and making agreements that human beings develop these capacities. Thus, it is through the creditor-debtor relationship and economic thinking that personhood takes form, through the first instances of instrumental reasoning and reasons-responsiveness.

Moreover, I argued that the creditor-debtor relationship introduced a new means for individuals to gain power and status within the community. Economic transactions and trade become a source of wealth that set some individuals apart from others. This provides a means for the individual to stand out (because successful in trade), by providing benefits to the community (through surplus goods). This allows for self-

evaluation that is not based on conformity with others (in order to avoid being perceived as suspicious and dangerous) and fosters a sense of comparative self-worth that aims at achieving status as an individual.

Nietzsche also appears to claim that the creditor-debtor relationship between individuals existed prior to any human community. This creates three problems. First, it is implausible to imagine individuals developing contractual arrangements to exchange goods on their own, and even more implausible to imagine this succeeding, absent some community to enforce those agreements. Second, it goes against contemporary anthropological evidence, making Nietzsche's simply false. Third, Nietzsche claims that human beings have always existed in communities, so this would make him guilty of inconsistency.

To resolve this issue I noted that the German Nietzsche uses for "societal associations and organizational forms" (*gesellschaftlichen Organisationformen und Verbänden*) is used to denote communities with political structures, while his use of "communal complexes" (*Gemeinschaft*) denotes communities organized according to blood relationships. Nietzsche never claims that the creditor-debtor relationship existed prior to communities of people who were blood relations, he claims that it existed prior to any politically organized society. This, I argued, is consistent with anthropological evidence of nomadic trading cultures existing long before any settled cities. The creditor-debtor relationship developed within the context of nomadic communities, where contracts were enforced and the creditor-debtor relationship was made possible.

Nietzsche appeals to human cruelty to explain several developments in his account. In the first political state, human aggression undergoes forced internalization, and this could provide an explanation for the development of human cruelty. But Nietzsche claims that the creditor's pleasure in causing suffering to the debtor explains how punishment provides repayment for a debt, and the creditor-debtor relationship existed prior to the state. This indicates that Nietzsche sees cruelty as a basic feature of human nature and not caused by the internalization of aggression. But human communities require cooperation to function and survive, and this does not seem possible if its members are motivated to make each other suffer. Moreover, it fails to explain how human beings developed the capacity to engage in altruistic behavior.

I responded to this problem by arguing that natural selection favored early human communities that developed aggressive instincts and a capacity for violence. Aggression and violence were necessary for the community to protect itself from predators and enemies, and to survive through hunting. But natural selection also favored communities who cooperated and protected each other. Thus, natural selection favored communities who directed aggression and cruelty outside of the community, and engaged in reciprocal and cooperative behavior with other members of the community. This explains how human beings developed instincts for cruelty, but also developed instincts for altruistic behavior.

The creditor-debtor relationship appears in three forms in GM II: between individual buyers and sellers, between the community and the individual, and between the

community and its ancestors. An issue of interpretation is to explain the order in which these relationships appeared. The fact that a community had to exist in order to enforce contracts between individuals, appears to indicate that the creditor-debtor relationship between the community and individual had to exist first. But Nietzsche claims that the economic thinking that developed between individual buyers and sellers “first transferred onto the coarsest and earliest communal complexes.”

I argued, however, that the creditor-debtor relationship is an early instance of interpretation and noted Nietzsche’s distinction between a thing or practice and its meaning. Interpretations are always applied to already existing things or practices. In the relationship between buyers and sellers, the existing practices were basic forms of trade and exchange between members of the community involving readily available goods. The creditor-debtor relationship becomes necessary when some members have surplus goods to exchange for the promise of future repayment. This transforms sellers into creditors and buyers into debtors, and makes keeping agreements a matter of obligation.

This structure is then interpreted onto the community and its members. The community is interpreted as a creditor that advances goods in its contractual relationship to the individual, who is interpreted as a debtor that owes repayment in the form of obeying the rules of the community. Similarly, and next in order, the practice of providing for the needs and desires of the ancestors is interpreted as repaying a debt for the benefits their efforts brought to the community.

These were the interpretive issues that were addressed in Chapter One. I now turn to the two most important features of Nietzsche's account of the creditor-debtor relationship for the purposes of this dissertation: the practice of punishment and the developments that take place in debt consciousness.

The purpose of punishment *within* the creditor-debtor relationship is to provide creditors repayment for unpaid debts. This purpose is only served, however, because the creditor enjoys causing the debtor to suffer, and takes this pleasure as compensation for his material loss. Punishment thus serves to anesthetize feelings of injury over the loss, through violent affect. The loss is not actually remediated through punishment, rather, the feeling of loss is muted through the satisfaction of causing suffering. Thus, the unquestioned assumption underlying the dynamics of the creditor-debtor relationship is that the human desire to cause suffering over an injury must be satisfied, and the creditor-debtor relationship justifies doing so by interpreting the desire to retaliate as punishment for a broken promise and unpaid debt.

The more important consequence of punishment in the creditor-debtor relationship is that it established a psychological connection between debts, as instances of broken obligations, and suffering. Once established, this becomes a connection between anything interpreted as a debt – such as owing the community obedience to customs, or owing the ancestors sacrifice – and the prospect of suffering for failure to repay that debt.

The interpretation of the community and the ancestors as creditors changes the nature of debt and debt consciousness. In the relationship to the community, because the debt

owed by the individual is directly proportional to the power of the community and the benefits it provides, debt consciousness grows. And because the requirement to repay one's debt through obedience to customs is an ongoing matter, debt consciousness becomes constant. Moreover, because the community has a monopoly on the advantages it provides and its rules are non-negotiable, this increases the feeling that repaying debt through obedience is a matter of utmost necessity. To break the rules of the community is to decline all of the advantages it provides, and in the earliest stages amounts to self-banishment.

The fact that the individual's relationship to the community is interpreted as contractual encourages the idea that the individual voluntarily entered into the relationship with the community, and that disobedience to customs is a broken promise. Here the creditor-debtor relationship first provides a justification for the community's already existing practice of banishing customs-breakers and later provides justification for other forms of punishment that replace banishment. In each case, however, the community's response to the criminal is understood as laying claim to the debt that is owed, by making the criminal suffer.

I also argued that the most important development that occurs in the creditor-debtor relationship to the community is that interpreting obedience to customs as the debt that is owed, strains the original material debt model. In the relationship between buyers and sellers, there is in principle some material compensation for the debt that would allow the debtor to avoid punishment. But when a broken rule – a crime – is understood as an

unpaid debt *per se* (independent of any harms or material losses caused), there is no material compensation that can be provided, and so the only means of repaying the debt is through suffering.

The creditor-debtor interpretation of the ancestors and community is even more strained and produces several important developments in debt consciousness. I argued that as the conception of the ancestors develops from powerful spirits to gods, the nature of their desires becomes more inscrutable and produces increasing suspicion that the community's sacrifices are inadequate. This is complicated by the fact that – as spiritual beings – the gods are not present to the community to make their satisfaction or dissatisfaction with these sacrifices known, which further increases doubt that the sacrifices have been adequate. This leaves determining the proper sacrifice to the fearful imagination of the community, leading to greater sacrifices and increased anxiety over punishment. Thus, through the relationship to the gods, debt consciousness is not only ongoing, but continues to escalate in intensity. This increases anxiety over punishment and increases the worry that the community is incapable of repaying its debt.

Because the gods are responsible for the community's existence and success, the contract between the individual and community extends to the gods. This results in a peculiar understanding of the contract and the nature of the gods. The contract, in this case, seems to be involuntary. The gods have already provided, and continue to provide the community with benefits and expect repayment for doing so. The debt is an obligation *imposed* on the community unilaterally by the gods. Thus, the relentless effort

to interpret every relationship in terms of creditors and debtors, confuses the difference between a creditor and a benefactor. Cultures that interpret the gods strictly as creditors remain caught in a religious model based on fear and punishment. Cultures that move outside of this model create the possibility for understanding the gods as providing benefits that are repaid with gratitude, rather than fear.

The relationship to the gods differs from the other forms of the creditor-debtor relationship in another remarkable way. In the other forms, suffering punishment imposed by an individual creditor or the community provides repayment for debt and relieves the burden of debt consciousness. But, while the community's success is interpreted as evidence of the gods' aid, misfortunes are not interpreted as the gods' punishment, but rather, as an indication of their diminished power. Thus, so long as the gods are thought to exist, debt consciousness only escalates and is never diminished. This makes anxiety over punishment even more unfocused and vague and increases further suspicion that the debt has not been repaid.

This indicates a community that sees the gods as only interested in repayment, and not interested in relieving debt, which leaves the community no way to address debt save through their own (increasingly doubted) efforts. This produces tension in the community's understanding of its rules and customs and marks an important turn in the development of debt consciousness that is significant for guilt. Since the gods established the community and its customs, obedience to customs is taken very seriously. Any act of disobedience threatens the entire community, so the commands of the gods take on

categorical force. But the community is convinced that the gods require more than obedience to customs. Sacrifice becomes the central means to repay debt. Thus, strict obedience to customs becomes a necessary, but insufficient condition for repaying the gods.

This is a significant sequence of developments in the history of morality and the feeling of guilt. Anxiety over punishment has turned religious obligations into matters of ethical concern. Ritual failures threaten the community, so are wrongs that warrant punishment. Through the moralization of religious practices, fulfilling one's religious obligations becomes a matter of living up to ideals and embodying particular virtues. This leads to the idea that failing to live up to an ideal is an ethical wrong that deserves punishment. This conflation of ethical obligations and aspirations to an ideal of human nature is a central feature of the underlying phenomenon that is interpreted as guilt. Failing to live up to an ideal is a personal matter and so produces feelings of personal unworthiness. Because this is often a private matter, punishment for this failure is left to the individual, and provides a pretext for self-punishment and feelings of worthlessness. Moreover, because the relationship to the gods produces the feeling that obedience is both imperative and inadequate to provide repayment, failure fosters a sense of futility and impotence that leads to resentment and the bad conscience.

Chapter Two presented the development of the bad conscience and guilt as the product of the internalization of the instincts, inescapable debt consciousness, resentment, and the ascetic ideal.

The origins of the bad conscience are found in the internalization of aggressive, animal instincts, which occurs when nomadic cultures are enslaved and forced into a politically organized community, and into conditions that deny external expression of those instincts. This produces what is known as the “animal bad conscience.” I argued that the conditions of “society and peace” are not sufficient to explain the development of resentment and the bad conscience. I claimed that the slaves, in addition to having an instinct for freedom, also had a long-engrained instinct to obedience. When aggression is denied external expression and seeks expression internally, it does so through denying expression to the other animal instincts, thus, becoming a will to power over the other instincts. This will to power is coordinated with the instinct to obedience and allows the slaves to adapt to their new conditions and become obedient to new rules and customs. This satisfied their will to power and so did not lead to resentment.

I argued that Nietzsche presents a general account of the development of resentment and the bad conscience that can be connected to actual historical events. On the general account, resentment is produced when the slaves have adapted to dominant noble culture, inherit the value distinction between “good and bad,” and adopt the gods of the nobles, thereby inheriting the burden of the unpaid debt to those gods. According to noble values, the slaves are irredeemably “bad” and incapable of living up to the ideals that would provide the gods with repayment. The slaves are burdened with a sense of debt consciousness and anxiety over punishment that they are impotent to avoid or ameliorate, which produces suffering. The impotent desire to locate and retaliate against the cause of

their suffering produces resentment. The slaves's suffering and underlying feelings of anxiety and worthlessness are interpreted by the ascetic priest as guilt before God, and punishment for sin. This redirects the slaves resentment towards the cause of their suffering, which is identified by the priest as their own animal instincts. This provides the slaves with a pretext to vent their resentment and pent-up aggression on themselves through self-inflicted cruelty and punishment, thereby satisfying the will to power and "saving the will."

In the historical account, provided in Chapter Three, I argue that the history of post-Exilic Judaism fills in the gap in Nietzsche's narrative. His presentation of the origins of the bad conscience begins with a vague and clearly pre-historical account of the subjugation of nomadic cultures by "some pack of blond beasts of prey," but it culminates with Christianity and the doctrine of the atonement. The historical events that connect to Christianity begin with the fall of Jerusalem, the destruction of the first temple, and the exile of the Jewish people to Babylon. These events are significant because they spell the end of any independent Jewish monarchy and the warrior culture that constituted it, and, through the destruction of the temple, the end of the sacrificial system.

The end of the sacrificial system creates two related problems. The first is a crisis of religious interpretation. Without the temple and sacrifice, Israel's relationship to God is severed. The surviving priests are forced to provide a reinterpretation of Israel's covenant with God in light of this loss to preserve themselves and Jewish identity. The second

problem is that the sacrificial system provided a way for the Jewish people to internalize their aggressive instincts. The Jewish people find themselves a subjugated people in a culture with new rules and customs and deprived of any way to satisfy their aggressive drives, and suffer from the animal bad conscience.

The priests reinterpret Israel's situation in exile as God's punishment for failure to meet God's covenant demands, now interpreted as matters of ritual purity and ethical stringency. They identify the cause of this failure in the base desires of the people (including especially the warriors kings). The priests address the loss of the temple by reinterpreting sacrifice as an internal matter of study, prayer, devotion to God, ritual and ethical purity, and living up to God's standard of holiness through self-denial.

Without an opposing warrior-noble culture, the priests' interpretation of "good and bad" as "pure and impure" rises to dominance. As the Jewish people accept this reinterpretation of Israel's history and religion they accept the distinction between "pure and impure" and accept the priests' ascetic interpretation of a holy God. In doing so they inherit the "debt" of living up to standards of purity and holiness that they are incapable of achieving, which leaves them under the sentence of God's judgment and anxious over divine punishment. The priest interprets the suffering they endure from the animal bad conscience and the burden of debt as the bad conscience. They are guilty of sin and impurity in the eyes of a holy God, and are being punished for their sin. I argued that this is why Nietzsche claims that the slave revolt in morality began with the Jews. It begins with the priestly rejection of warrior-noble morality.

The subsequent history of Israel is a series of conquests by other foreign powers ending with the Roman occupation. According to priestly morality this is evidence that Israel has not fulfilled its covenant requirements through purity. The Jewish people are caught in a religious value system that makes them responsible for an unfulfilled covenant and destiny, because of their sinfulness, which produces resentment and suffering. The slave revolt culminates with Jesus' radically egalitarian rejection of all noble morality (priestly and warrior) and the doctrine of the atonement.

Through the interpretation provided by the ascetic ideal, guilt and duty are "moralized" by making human nature itself – in particular its animal instincts – the source of the failure to live up to God's ideals. And through the influence of the creditor-debtor relationship, failure to live up to ideals is interpreted as a moral wrong that deserves punishment. Moralization escalates as the source of guilt is not only located in the individual, but is traced back to our human ancestors, the natural world, and existence itself. This increases anxiety over guilt and punishment, as every aspect of human existence is impugned as a source of sin.

Nietzsche claims that the doctrine of the atonement provides temporary relief from this escalation of guilt. Anxiety is lifted through the idea that God took on the punishment for guilt. But I argued that this relief is only temporary because it affirms that human beings are guilty, and dramatically reinforces the connection between guilt and suffering. In doing so, it ultimately leaves punishment unaddressed by leaving to the "guilty" individual to address their own guilt through self-punishment. Thus, the atonement does

not eliminate guilt and only relocates the source of punishment to the individual. Because this satisfies the individual's will to power, the individual is motivated to pursue it, but because punishment only anesthetizes suffering temporarily, and because the individual becomes stronger and better able to endure suffering through suffering, this results in an ever-increasing dynamic of seeking sources of guilt as pretexts for self-punishment.

Nietzsche expresses horror over the idea that a creditor would sacrifice himself "out of love" for a debtor. I argued that this is a reflection of the confusion between "creditor" and "benefactor" (or "lover") that is produced by the creditor-debtor relationship. If God is the benefactor, or lover, of humankind then the benefits bestowed on humans are a gift and not a loan, so there is no debt. And if God is a creditor that demands repayment, then the idea of self-sacrifice out of love is mistaken. One cannot be both creditor and lover.

Nietzsche claims that, through the moralization of guilt and duty, feelings of guilt do not decrease with waning belief in God, as one might plausibly expect. I argued that this is because moralization identifies the animal aspects of human nature as the source of wrongdoing and understands failures to attain an ideal as wrongs deserving of punishment. As these ideas are pushed back into the human conscience, they survive atheism by being reinterpreted in secular terms. The source of wrongdoing becomes "selfishness," or "egoism," or "irrationality," and the failure to live up to ideals of selflessness, altruism, or perfect rationality produce as much guilt as the concepts of sin in relation to holiness.

According to my interpretation the development of the bad conscience occurs through post-Exilic Judaism and is produced when sacrifice is no longer available as a means to internalize aggression. Contrasting this situation with the ancient Greeks shows how they were able to avoid the development of the bad conscience. I argued that one factor was the nature of the ancient Greek gods, who represented idealized versions of human nature in both its good and bad qualities. Through this, the ideals that the ancient Greeks aspired to were both attainable and did not condemn the animal aspects of human nature. Moreover, these gods were responsible for the evils deeds that humans commit, by “beguiling” them into evil. Thus, the Greeks did not feel pervasive guilt, because of some aspect of human nature. More importantly, I argued that sacrifice offered the Greeks a way to internalize their aggression, while at the same time recognizing that the cruelty that was necessary for a genuine sacrifice was the same cruelty exhibited by the gods and their own nobility, which offered them a way of affirming themselves as noble and good.

The results of contrasting ancient Greek culture with Judeo-Christian culture supports two important conclusions for this dissertation. The first is that the bad conscience is a perversion of the conscience and sovereign individuality. According to my interpretation, the bad conscience in GM II is the product of a specific, ascetic-religious interpretation of debt consciousness that took place through Judaism and Christianity. This implies that Nietzsche’s primary concern is with the feeling of pervasive guilt that manifests itself as anxiety over a sense of culpable personal failure and worthlessness that warrants self-punishment and suffering. This feeling of guilt was only possible because the slaves

inherited noble values and adopted the nobles' gods. This produced the sense that they were "bad" and incapable of living up to ideals that would allow them to repay their debt to the gods, which produced anxiety over unavoidable debt and punishment, which is interpreted by the priest as the bad conscience. But to accept the judgment that they are "bad" and fall short of noble ideals the slaves either had to already be, or had to become along the way, sovereign individuals. The slaves deny their own natures as falling short of the ascetic ideal, and seek to self-regulate their behavior in light of that ideal. Thus, they are capable of valuing and acting on values, rather than momentary affects and desires, and are sovereign individuals.

But the slaves attained sovereign individuality through the interpretation of their suffering as guilt. The slaves deny their animal natures in light of the ascetic ideal, because it provides meaning to their anxiety, feelings of worthlessness, and suffering, and provides a goal for them to will. But they are motivated to do so because this interpretation has provided a target for their resentment, a means for discharging their aggression and cruelty, and an ideal that guarantees that the pretext to engage in self-torment will never be exhausted. Thus, the slaves' sovereign individuality is founded upon their own self-loathing and internalized cruelty and only amounts to a project of satisfying cruelty and self-anesthetization through violent affect, that never addresses the actual cause of their suffering. And as long as this project involves the rejection of their own natures, this tremendous expenditure of energy remains a vicious circle that produces no positive growth toward an attainable ideal. Thus, sovereign individuality has

been perverted into an insulated, self-reinforcing dynamic of suffering and cruelty for its own sake.

The central thesis of this dissertation is that Nietzsche's call to eliminate the concepts of guilt and punishment is plausible and to defend him from the objection that doing so would produce undesirable consequences. In the case at hand, the objection is that eliminating the feeling of guilt would entail that human beings never feel pain over wrongdoing. If this were the case, then Nietzsche's view would entail diminished concern for the well-being and standing of other persons which is antithetical to a genuine and robust ethical life. I argued that this consequence is not entailed by Nietzsche's rejection of guilt. Nietzsche rejects the bad conscience and the feeling of pervasive guilt, but he does not reject the conscience itself, and the existence of a conscience entails the capacity to feel pain over wrong actions. I call this "episodic guilt" (other interpreters refer to it as "locally reactive guilt") and claim that Nietzsche's only concern with episodic guilt is when it is connected to pervasive guilt and is used as a pretext to promote that feeling and engage in self-punishment. This claim was supported by examining the difference between the conscience and its distortion in the bad conscience.

The conscience was produced through the morality of custom and its use of pain to instill a feeling of "obligation" in human beings to follow the rules of the community. When this connection becomes a feature of a person's psychology, when it is internalized, and "forgotten" (through the active faculty of suppression), and made instinctive, along with the rules of the community, it remains as a feeling of pain over "ethical" failure, and

the person has a conscience, i.e. an internalized mechanism of self-sanctioning that allows self-regulated behavior. This is what frees the sovereign individual from the morality of custom and the need for external supervision and the threat of punishment. Thus, the conscience serves the essential purpose of guiding one's actions in a community, in light of norms that articulate the concerns, expectations, and "rights" of other persons.

It is highly implausible that Nietzsche rejects the conscience. The Greeks were able to build the civilization and culture that Nietzsche clearly admires by employing their cruelty to deny the expression of instincts that threatened to undermine that project. They sublimated their aggression towards building an ideal culture. Thus, the ancient Greeks were sovereign individuals with consciences, and Nietzsche presents them as *the* healthy contrast to the development of the bad conscience that took place through Judaism and Christianity. To accept the conscience as valuable is to accept the internalized connection between obligations and suffering as necessary for the conscience, and this is to accept the idea of being pained over ethical failure, i.e. "episodic guilt," which makes self-regulated behavior in light of norms and values possible.

The bad conscience presumes the existence of a conscience that has become consumed with thoughts of personal and moral inadequacy. It is a perversion of the conscience that employs the connection between ethical failure and suffering as a means to turn the conscience's healthy response to wrongdoing into an instrument of self-torture. Instances of ethical wrongdoing are taken as evidence of one's incorrigibly sinful

nature and a justification to punish oneself for one's ethical inadequacy. Thus, the bad conscience – the feeling of pervasive guilt – serves the purpose of satisfying the individual's will to power and anesthetizing suffering. As such, it is an entirely self-oriented, and self-serving dynamic that evidences no essential concern for the wrongs done to others, *qua* wrongs to others.

To eliminate “episodic guilt” is to eliminate the conscience and sovereign individuality, whereas to eliminate pervasive guilt is to eliminate the bad conscience and restore the conscience to its healthy functioning. Moreover, to eliminate pervasive guilt and the bad conscience is to eliminate a psychological dynamic that distracts the individual from the effect of wrongdoing on others. Thus, to eliminate guilt, in Nietzsche's sense, is to eliminate pervasive guilt and a mechanism of self-torture, but inasmuch as Nietzsche regards the conscience as valuable, this does not entail eliminating “episodic guilt” and feeling pain over one's wrongdoing, but rather, restores this feeling to its proper function: guiding one's actions in accordance with ethical norms in their relationship to other persons.

I also argue that Nietzsche's call to eliminate punishment is plausible and defend the claim that this does not entail undesirable consequences. In the case of punishment, the worry is that “eliminating punishment” means leaving crime unaddressed and allowing criminals to run free. This is often thought to be implied by Nietzsche's claim that he can imagine a society with such a consciousness of its own power that it allows itself the luxury of leaving its criminals unpunished. If Nietzsche's view entails this result, then it

is objectionable from both a moral and practical standpoint. It is morally objectionable for a community to fail to protect its members by leaving crime unaddressed. It is practically objectionable, because leaving crime unaddressed would undermine the community's stability and either lead to its dissolution and decay, or prompt a return to more severe forms of punishment.

Examination of the creditor-debtor relationship revealed that inflicting suffering on debtors was an interpretation of the already existing community practice of banishing customs-breakers, which in turn was rooted in the community's treatment of defeated war-time enemies. The creditor-debtor relationship allows creditors to inflict suffering on insolvent debtors as a means of compensation for material losses. In this way, instinctive retaliation over injury is turned into punishment by providing it with a justification and purpose. Retaliation is justified because the debtor broke a voluntary agreement to provide repayment, resulting in injury, and the debtor's suffering accomplishes the aim of providing the creditor with satisfaction through causing suffering. Suffering is now justified as the means of addressing a perceived injury, and this is punishment in its essence. The results that punishment has been expected to produce are numerous. They have changed over time as punishment has been reinterpreted to have different ends, but the means to those ends have always remained the same: intentionally inflicted suffering. Thus, punishment is the justification of intentionally inflicted suffering.

I argued that this reveals the motivation for punishment to be rooted in the injured party's anger and desire to retaliate against the cause of injury. When the creditor-debtor

relationship is interpreted onto the relationship between the community and individual, breaking customs is interpreted as a failure to repay the debt of obedience, and the community's harsh treatment of the the customs-breaker is justified as recouping that debt, initially through banishment, but later replaced by other forms of punishment. In either case, however, the community's motivation to punish is also the desire for angry retaliation that is satisfied through the customs-breaker's suffering. Thus, an economic interpretation of the community is responsible for establishing and reinforcing the idea that the criminal deserves to suffer, and must suffer for his crime, but this is an interpretation that masks the underlying motivation to enjoy the pleasure of making suffer.

The desire to retaliate over an injury and cause suffering in return is the desire to take revenge. Nietzsche identifies this motivation as resentment, and I argued that his views on justice show us that the stability of a community is threatened by both crime and the resentment of those who feel injured and wronged. This is why he claims that those who administer justice in a community work in opposition to the demands of resentment. Nietzsche describes resentment as falsifying its target by making it into a monstrous caricature, and as only committed to the perspective of the injured party. I argued that this is because resentment seeks a pretext to vent itself fully on some target that it can blame for the totality of its suffering. But since the victim suffers from causes other than the injury caused by the crime, resentment seeks to create a target that justifies the full brunt of the victim's pent-up aggression. Left unaddressed, resentment

leads to extreme responses to crime that threaten to escalate and threaten the stability of the community.

I take this to be more than sufficient evidence to call into question the morality of punishment. If there is any essential difference between the motivation of the criminal and those who punish, it is that the criminal disregards the suffering of his victims in order to achieve his goals, while the punishers directly aim at causing suffering and take pleasure in doing so. At best the motivations of the criminal and the punishers are symmetrical, reflecting the same desires, and the same pleasure in causing suffering. It is difficult to reconcile punishment as morally justified, when it shares the same motivations as the criminal. At worst, the criminal's motivation is better than the punishers, since it may only show ignorance, carelessness, or indifference to suffering without, however, having the direct aim of causing suffering.

Nietzsche provides further evidence, however, to consider the elimination of punishment as in the service of morality. I argued that one of the effects of punishment is to anesthetize the feelings of the punisher through violent affect. Causing suffering discharges aggression and produces a feeling of power and pleasure that "compensates" by being a more powerful stimulus than the feeling of loss or injury. Beyond providing pleasure through cruelty, punishment anesthetizes any legitimate feelings of moral resentment, outrage, or protest. Thus, punishment distracts the victim from the aim of moral protest which is recognition, either from the criminal or the community, or both, of the victim's moral standing and the lack of respect shown for that standing through the

crime. Punishment provides the victim with the pleasure of the criminal's suffering, which is a cheap and easy substitute for moral regard.

The foregoing constitutes the core of my critique of David Hershenov's debt and revenge model of punishment. I argued that this model serves to justify the victim's immoral motivation to take pleasure in the criminal's suffering. In the process it distracts both the criminal and the victim from the wrong that was done and the real debt that is owed, which respect for moral standing. Punishment only serves to direct the criminal's attention to his own suffering and away from the nature of the wrong done. Especially when the criminal recognizes the victim's motivation as the desire to enjoy his suffering. This teaches the criminal that causing suffering is not wrong in and of itself, but must be based on a suitable justification. Punishment only distracts the victim by anesthetizing the suffering caused by resentment, thereby undermining the aim of moral protest and the demand for recognition of moral standing. Moreover, as a justification for punishment, the revenge model also justifies and encourages the feeling of resentment and thereby works against the aim of justice, which is to eliminate the influence of resentment in the community.

The problem with any justification of punishment, I argue, is the assumption that the appropriate response to crime is to intentionally inflict suffering on the criminal. This is the core definition of punishment that goes largely unquestioned. And it is this assumption – that the means of addressing crime is punishment, and that punishment is “hard treatment” of the criminal – that leads to crises of justification and the view that, if

punishment as hard treatment is not justified, then incarceration is not justified, and that the rejection of punishment entails leaving crime unaddressed.

Nietzsche has provided us with the means for seeing that this is not the case. Punishment is an interpretation of the pre-existing practice of addressing “crime” in early human communities. The assumption was that the anger of the community would lead to the “criminal’s” suffering, and punishment provided a justification for this practice. But to reject the concept of punishment is not to reject the idea that crime must go unaddressed, it is to reject the interpretation that crime must be addressed by causing suffering. When Nietzsche claims that punishment should be eliminated, he claims that it should be eliminated from our social sanctions. The distinction between punishment and a sanction resolves the worry about Nietzsche’s view. A sanction is a community’s way of addressing crime. A sanction could be interpreted as punishment, but it need not be so interpreted.

I argued for interpreting sanctions for crime on a containment and reform model. On this model, one aim is to protect the community by isolating a criminal threat through incarceration. The practice of incarceration, on this model, does not directly aim at causing suffering; it only aims to protect the community from further harm. Since incarceration does not involve intentional hard treatment, this is the only justification that it requires. The criminal may suffer indirectly through loss of freedom, but this is not the aim of incarceration, and is one of the reasons why the containment model must include efforts to reform or rehabilitate the criminal. Those who are motivated by moral

considerations can recognize that reform of the criminal is a way of showing respect for moral standing, and may provide the opportunity for the criminal to be restored to the community as a genuine moral agent who can work to make good for his crime. There are also pressing practical considerations that support the need for reform and rehabilitation that do not appeal to moral concern for the criminal. In most cases, incarcerating the criminal for life is both practically infeasible and morally objectionable. This implies that many criminals will be returned to the community. If the purpose of justice is to protect the community from crime, then those who administer justice must take this fact into account, and recognize that moral reform, where possible, is necessary to provide protection to the community.

The history of punishment is the history of the economics of suffering and cruelty. Nietzsche's genealogy of the creditor-debtor relationship reveals how this began. Our early human ancestors interpreted their primitive responses to injuries and threats in terms of the only developed concepts available to them: economics. Once injury and suffering are identified as a source of debt, this entails that injury and suffering are the currency in which the debt must be repaid. The concept and practice of punishment is predicated on this interpretation. At its best punishment is a compromise with the anger and resentment of the injured party. At its worst, it is a justification for indulging in the pleasures of cruelty. Nietzsche's general distinction between practice and interpretation, and the specific distinction between sanctions and their interpretation as punishment, reveals how punishment might be plausibly brought to an end.

Finally, I argued that the moral interpretation of guilt has influenced our interpretation of other forms of anxiety in unhelpful ways. Because of the prevalent influence of morality on human psychology we have a tendency to explain experiences with a particular phenomenal feeling in moral terms. This is because human beings were first led to identify originally non-moral feelings in moral terms, and this tendency spread and grew through the influence of Judeo-Christian morality on human development in Western culture. This tendency survived the decline of religious belief and secularization through the moralization of guilt. This tendency was illustrated in David Velleman's explanation of "survivor guilt." I argued that, because he begins with a general definition of guilt that is drawn from an understanding of moral guilt, he is led into a fruitless and misguided attempt to articulate a non-moral normative standard for surviving disasters that will justify the victim's envious-resentment of the survivor's undeserved good, and hence explain "survivor guilt" as rational.

I proposed that Velleman's notion of "normative vulnerability" be replaced with the more basic phenomenon of "relational vulnerability," which explains the development and importance of normative standards to human beings in a variety of contexts, and also explains experiences such as "survivor guilt" in a manner that does not promote unhelpful moral resonances that lead to feelings of deserved suffering and worthlessness.

Guilt, like punishment, was shown to be rooted in the human impulse to cruelty, in this case self-cruelty, which has become the prevalent means for human beings to vent aggression and satisfy the will to power. Given the influence of morality on human

psychological development, and its continuing influence on human thoughts and feelings, interpreting forms of anxiety automatically in terms of guilt is far too common and easy, and should be resisted. Eliminating the tendency to think in terms of guilt would allow us to identify the basic relational vulnerability that is the source of anxiety and would allow us to clarify both our normative standards and our relationship to them, hopefully in a way that is responsive to human vulnerability and need.

Bibliography

Bright, J. (1972). *A History of Israel* (3rd ed.). Philadelphia: The Westminster Press.

Clark, M., & Dudrick, D. (2012). *The Soul of Nietzsche's Beyond Good and Evil*. New York: Cambridge University Press.

Conway, D. (2008). *Nietzsche's On the Genealogy of Morals*. New York: Continuum International Publishing Group.

Hershenov, D. B. (1999). Restitution and Revenge. *Journal of Philosophy*, 96(2), 79-94.

Luther, M. (1962). *Martin Luther: Selections from his Writings* (J. Dillenberger Ed.). New York: Doubleday.

Morris, H. (1968). Persons and Punishment. *Monist*, 52, 475-501.

Morris, H. (1971). Guilt and Punishment. *Personalist*, 52(305-321).

Morris, H. (1981). A Paternalistic Theory of Punishment. *American Philosophical Quarterly*, 18, 139-152.

Murphy, J. G., & Hampton, J. (1988). *Forgiveness and Mercy* (J. Coleman Ed.). Cambridge: Cambridge University Press.

Nietzsche, F. (1990). *Twilight of the Idols / The Anti-Christ* (R. J. Hollingdale, Trans.). New York: Penguin Books.

Nietzsche, F. (1992). *Ecce Homo* (R. J. Hollingdale, Trans.). New York: Penguin Books.

Nietzsche, F. (1998). *On the Genealogy of Morality* (M. Clark & A. J. Swensen, Trans.). Indianapolis: Hackett Publishing Company.

- Nietzsche, F. (1999). *Daybreak* (R. J. Hollingdale, Trans. M. Clark & B. Leiter Eds.).
New York: Cambridge University Press.
- Nietzsche, F. (2001). *The Gay Science* (J. Nauckhoff & A. Del Caro, Trans. B. Williams
Ed.). New York: Cambridge University Press.
- Nietzsche, F. (2003). *Beyond Good and Evil* (R. J. Hollingdale, Trans.). New York:
Penguin Books.
- Nietzsche, F. (2005). *Thus Spake Zarathustra* (G. Parkes, Trans.). New York: Oxford
University Press.
- Reginster, B. (2011). The Genealogy of Guilt. In S. May (Ed.), *Nietzsche's On the
Genealogy of Morality a Critical Guide*. New York: Cambridge University Press.
- Reginster, B. (2018). What is the Structure of Genealogy of Morality II? *Inquiry*, 61(1),
1-20.
- Ridley, A. (1998). *Nietzsche's Conscience: Six Character Studies from the "Genealogy"*.
Ithaca, NY: Cornell University Press.
- Ridley, A. (2005). Guilt Before God, or God Before Guilt? The Second Essay of
Nietzsche's Genealogy. *Journal of Nietzsche Studies*, 29(Spring), 35-45.
- Risse, M. (2001). The Second Treatise in On the Genealogy of Morality: Nietzsche on the
Origin of the Bad Conscience. *European Journal of Philosophy*, 9(1), 55-81.
- Snelson, A. (2016). The History, Origin, and Meaning of Nietzsche's Slave Revolt in
Morality. *Inquiry*, 60(1-2), 1-30.

Velleman, D. J. (2006). "Don't Worry, Feel Guilty" *Self to Self: Selected Essays*. New York: Cambridge University Press.