10

Vengeful Thinking and Moral Epistemology

Neil Sinhababu

Daybreak, the book in which Nietzsche's "campaign against *morality* begins" (*Ecce Homo*, D:1), itself begins with the following aphorism:

> *Supplemental rationality*—All things that live long are gradually so saturated with reason that their origin in unreason thereby becomes improbable. Does not almost every precise history of an origination impress our feelings as paradoxical and wantonly offensive? Does the good historian not, at bottom, constantly *contradict*? (1)

Six years later, in *On the Genealogy of Morality*, Nietzsche traces the historical development of contemporary moral beliefs from their "origin in unreason." While the *Genealogy* may have impressed the feelings of many readers as "wantonly offensive," its import for moral theory has not been fully understood, and the focus of this essay will be to make clear the significance of histories like Nietzsche's for contemporary debates about whether belief in moral properties is justified. One might worry that arguments against believing something whose premises involve the ways that the belief originated will commit the Genetic Fallacy, mistakenly claiming direct logical connections between the circumstances of the belief's origin and its truth. There are, however, good arguments connecting the origins of beliefs to their justificatory status, and I will present two such arguments in this essay.

The first essay of the *Genealogy* describes the way that the vengeful fury of oppressed classes in ancient Rome caused them to invert the value system of their rulers and embrace slave morality. Nietzsche's account does not rely on moral properties to do any explanatory work, and it presents the slaves' mechanisms of belief-formation

as ones that are unreliable in generating true belief. If this account (or an account that is like it in the relevant respects) is correct, it will support a pair of arguments that moral anti-realists could use against non-reductive naturalistic moral realism. According to the Explanatory Argument, we should reject moral properties because the best explanation of our observations makes no reference to them. Gilbert Harman makes this argument in *The Nature of Morality*, and Brian Leiter deals with realist counterarguments to it in "Moral Facts and Best Explanations". According to the Unreliability Argument, processes of belief-formation that are unreliable in getting us to the truth are responsible for our beliefs in moral properties. As many plausible theories of epistemic justification claim, we have little reason to hang on to beliefs that we know were caused by unreliable processes. Therefore, we should give up the beliefs about moral properties that we know to be caused by these unreliable processes. This argument for anti-realism has not attracted as much attention, and it will be a major focus of this essay to develop it.

First I will go over Nietzsche's account of the origins of our moral practices in the first essay of the *Genealogy*. Then I will address the disputes about moral realism that the Explanatory Argument and the Unreliability Argument will enter into, and spell out these two arguments. I will show how Nietzsche's account advances the arguments by giving us reason to accept their premises.

According to Nietzsche's account in the *Genealogy*, the contemporary Western system of moral values was born thousands of years ago in ancient Rome when oppressed slaves embraced a moral system that was the reverse of their noble oppressors' values. Both systems of moral values were formed under conditions of intense emotion, not out of cold calculations of prudence or after sober reflection on what is valuable. The former set of values arose from the pride of the nobles, who first "ranked themselves and their doings as good" (GM I:2). Proud of themselves and their deeds, nobles' evaluations of other people reflected the degree of similarity between these people and themselves. Actions that expressed a powerful, brave, truthful, honorable character similar to that of the nobles themselves were admired, while weak, cowardly, dishonest, and base actions that expressed an ignoble character met with disapprobation. In *Beyond Good and Evil*, Nietzsche describes this kind of valuing and the emotional state that drives it: "Everything it knows as part of itself it honors: such a morality is self-glorification. In the foreground there is the feeling of fullness, of power that seeks to overflow" (260). The nobles' beliefs about what is valuable arise out of emotion rather than clearheaded reflection on what would serve them best: "The viewpoint of utility is as foreign and inappropriate as possible, especially in relation to so hot an outpouring of the highest rank-ordering, rank-distinguishing value judgments: for here feeling has arrived at an opposite of that low degree of warmth presupposed by every calculating prudence, every assessment of utility" (GM I:2). Nietzsche describes the mental states that give

rise to all systems of values in *Thus Spoke Zarathustra*: "Good and evil have always been created by lovers and creators. The fire of love glows in the names of all the virtues, and the fire of wrath" (Z I: "On the Thousand and One Goals").

It is the fire of wrath that glows brighter in slave morality. The historical preconditions for the formation of slave morality involve the oppression of slaves at the hands of their masters. The slaves are infuriated at their mistreatment and at the miserable conditions of their lives, and since they are powerless to avenge their sufferings, their fury builds inside them. Their treatment by the nobles brings about the emotional state that Nietzsche calls *ressentiment*—a vengeful hatred of those whom they blame for their sufferings. While nobles are not immune to *ressentiment*, their power enables them to act on their hatreds and discharge their *ressentiment* quickly (GM I:10). But the slaves' powerlessness and inability to act causes their *ressentiment* to build up inside them until it begins to reshape their beliefs and values. The psychological effects of *ressentiment* play a major role in explaining how the slave morality arises—"the slave revolt in morality begins when *ressentiment* becomes creative and gives birth to values" (GM I:10).

How does *ressentiment* give birth to values and bring the slave morality into being? To understand this process, it is important to imagine how people full of *ressentiment* would see their oppressors—that their attitudes towards the oppressors would arise "out of the brewing cauldron of unsatiated hate" (GM I:11). Unsatiated hate, like other intense emotional states that include unsatisfied desire, can affect belief-formation.[1] This is nothing strange—in fact, it happens quite often in ordinary cases of wishful thinking. In cases of wishful thinking, an agent's desiring that *p* causes her to believe that *p*. Wishful thinking seems to play a key role in determining the slaves' beliefs about a number of topics. Since the wish that drives the slaves' belief-formation is a wish for revenge against their masters, driven by their unsatiated hate, "vengeful thinking" would be a more specific name for this phenomenon.

FN:1

As is the case with all varieties of wishful thinking, vengeful thinking disposes the slaves to believe that the states of affairs which they desire will come to pass. The more satisfying a kind of revenge against the nobles would be, the more powerfully vengeful thinking will dispose them to believe in it. As is generally the case in wishful thinking,

[1] There is debate about what kind of anti-realist Nietzsche is. I read him as an error theorist about existing moral discourse. This would justify the attribution of a view to him where moral dispositions are beliefs rather than, say, desires or norm-acceptances. Consider *Daybreak* 103: "it is errors which, as the basis of all moral judgments, impel men to their moral actions. This is my point of view . . . Thus I deny morality as I deny alchemy, that is, I deny their premises: but I do not deny that there have been alchemists who believed in these premises and acted in accordance with them." While it's possible that non-cognitivism would have appealed to Nietzsche, it is anachronistic to ascribe such an innovative view about semantics to him. Error theory, by contrast, has been around at least since Parmenides' claim that all talk of change was systematically mistaken.

the slaves acquire the belief that their revenge fantasies will come true, even when they lack sufficient evidence to justify this belief. The belief that they have the property of moral goodness while the nobles have the property of moral badness arises because of their desire for revenge. This set of beliefs about moral properties is satisfying to them in several ways. For one thing, it allows the slaves to regard themselves as fully justified in hating the nobles, as one is justified in hating evil people. For another thing, it plays a key role in explaining why the afterlife will be pleasant for them and terrible for the nobles. Since they are good and the nobles are evil, God will reward them while punishing the nobles.

The beliefs about the afterlife that result from vengeful thinking are presented in the long quotation from the early church father Tertullian, who vividly describes the torments that the enemies of Christianity will endure in hell when Judgment Day arrives. He describes the provincial governor and other "persecutors of the name of the Lord, being liquefied by flames fiercer than those with which they themselves raged against the Christians!" (GM I:15). Great tragedians, athletes, and philosophers who teach doctrines contrary to Christianity are all described roasting in hell. An important part of Tertullian's vision is that Christians will be able to watch from heaven and enjoy the suffering of their enemies. In fact, part of Tertullian's purpose in describing these torments is to convince other early Christians that pleasures far exceeding those of the public spectacles await them on the Day of Judgment. He exults in the thought of the pagans' torment: "What variety of sights then! What should I admire! What should I laugh at! In which should I exult, seeing so many and great kings . . . now groaning in deepest darkness!" (GM I:15).

There can be no denying the significance of vengeful thinking in causing someone to accept such a conception of the afterlife. Tertullian's intense desire for the nobles to suffer causes him to believe that they will have an afterlife of fiery torment. And his desire to have the fullest possible appreciation of their suffering causes him to believe that he will get a chance to see this suffering when he is in heaven and they are burning. Nietzsche highlights the role of vengeful thinking in proposing an inscription for the gate to heaven. Parodying the words on the gate to Dante's hell, Nietzsche remarks that "I, too, was created by eternal *hate*" would be an apt inscription on the gate to a heaven where Christians rejoice in the torment of their oppressors, "assuming that a truth may stand above the gate to a lie!" (I:15).

In this case, it is clear that the beliefs of the slaves are wildly in error. Even most theists who believe in an afterlife will regard Tertullian's beliefs about the nature of heaven and hell as false. Contemporary theistic accounts generally do not portray the inhabitants of heaven exulting in the miseries of the damned. I take this to be a powerful illustration of how vengeful thinking can lead one into error. Certainly, not all instances of vengeful or wishful thinking involve such absurd beliefs. Sometimes, these forms of belief-formation even give rise to true beliefs. But

I take it as uncontroversial that wishful thinking is an unreliable process of belief formation, as is the special case of vengeful thinking, and that the slaves' beliefs about the afterlife are an illustration of how badly awry vengeful belief-formation can go.

One of the most significant effects of *ressentiment* on the slaves—and the effect that the first essay describes in greatest detail—concerns the particular system of values that they adopt. In their hatred of the nobles and their "opposition to the aristocratic value equation (good = noble = powerful = beautiful = happy = beloved of God)" the slaves invert the noble system of values (GM I:7). Nietzsche presents the view about values that they come to hold:

> the miserable alone are good, the suffering, deprived, sick, ugly, are also the only pious, the only blessed in God, for them alone is there blessedness—whereas you, you noble and powerful ones, you are in all eternity the evil, the cruel, the lustful, the insatiable, the godless, you will eternally be the wretched, accursed, and damned! (GM I:7)

All the features of the nobles—especially the ones that were involved in the infliction of harm on the slaves—are regarded as evil. Nietzsche claims that the slaves' belief in the evil of the nobles arises first, and that the slaves' belief in their own goodness arises as a response. He invites us to "imagine 'the enemy' as the human being of *ressentiment* conceives of him—and precisely here is his deed, his creation: he has conceived of 'the evil enemy,' 'the evil one,' and this indeed as the basic concept, starting from which he now thinks up, as reaction and counterpart, a good one—himself!" (GM I:7). This is the process by which the values of slave morality—both the negative values associated with the nobles and the positive values opposite the nobles—are formed. Furthermore, it is the process by which the slaves come to make the distinctive set of moral observations associated with slave morality. In regarding the nobles as evil, they observe that the property of being evil exists in the world (specifically, in the part of the world that is the nobles). In regarding themselves as good, they observe that the property of being good exists in the world (namely, in the part of the world that is themselves). In addition to supporting a picture of the afterlife where their vengeance is carried out, this set of moral beliefs allows the slaves to see themselves as blessed creatures loved by God, while the nobles are hated by God as well as damned.

Here one might wonder about the justificatory status of the nobles' own moral beliefs. Their process of moral belief-formation is similarly driven by powerful emotions. Does Nietzsche think their beliefs are subject to the same criticisms? And if so, why does Nietzsche express more positive attitudes towards the noble morality than towards the slave morality? On my interpretation, Nietzsche regards the nobles' moral beliefs as unjustified and false as well. He describes how the noble mode of valuation sometimes "lays a hand on reality and sins against it" and "forms a wrong

idea of the sphere it holds in contempt" (GM I:10). But Nietzsche allows that one can rise to a fairly high level of human excellence without much epistemic virtue, and epistemic virtue is not the justification for his positive attitude towards the nobles. Nietzsche approves of the noble morality not because it is true or because the nobles are epistemically justified in accepting it, but because it promotes the active, proud, strong-willed lifestyle that the nobles enjoy.[2]

FN:2

I should deal with an interpretative issue here. Nietzsche occasionally uses language that suggests intentional, conscious action in describing how the slaves form their new set of values. He says that the slaves perform "an act of spiritual revenge" (GM I:7) that comes out of "vengeful cunning" and exhibits a "prudence of the lowest order" (GM I:13) when they "*fabricate ideals* on earth" (GM I:14). On this reading, one is almost tempted to add to the beginning of section 14 a scene where one of the priests in the dark underground workplace says, "Hey guys! Let's just get everyone to reverse all the nobles' values, and we'll have our revenge!" I have resisted this interpretation[3] and treated the slaves' inversion of values as the outcome of a passive and mostly subconscious process—vengeful thinking—and not as the intended outcome of an action. (Not much will hang on this regarding the larger argument which I will present later about the justificatory status of our moral beliefs—one need not invoke moral properties to explain a conscious and intentional inversion of values, and it would be quite surprising if the beliefs formed by this process were reliably caused.) One reason I have interpreted the inversion of values as passive and subconscious is that Nietzsche's account makes much less sense to me if the inversion of values is achieved by a conscious intentional action. One cannot change one's beliefs merely by desiring some state of affairs, believing that a belief-change will bring about the desired state, and performing the necessary belief-change through an act of will. Unfortunately for anyone who wants to make Pascal's Wager, beliefs do not change in this way. Even those who reject the assumption that moral values are beliefs will probably agree that values (whatever kinds of mental states they may be) do not change that way either. One can act because one believes that an action will bring about the desired vengeance, but one cannot simply swap the values one holds for another set of values as a means of inflicting vengeance. Powerful emotions like *ressentiment* can reshape one's values or cause one to regard new things as valuable, but intentional action is not the process by which a change in values occurs.

FN:3

[2] I cannot properly argue for this claim here. But let me point to BGE 4, where Nietzsche says that "the falseness of a judgment is not necessarily an objection to a judgment . . . The question is to what extent it is life-promoting, life-preserving, species-preserving, perhaps even species-cultivating." I contend that when Nietzsche speaks approvingly of noble morality, it is on these grounds.

[3] In Chapter 5 of this volume, Jay Wallace also criticizes this interpretation of the Genealogy, which he calls the "strategic interpretation." I am sympathetic to his criticisms, which differ in some respects from mine.

In section 13, Nietzsche twice describes as "self-deception" the process by which the slaves come to see merit in their weakness. Self-deception is not a process in which one forms a conscious intention to believe something and carries that intention out by an act of will. It is a subconscious process in which discomfort with the belief best supported by the evidence causes one to avoid accepting it or seriously considering the evidence for it. I consider wishful thinking to be a kind of self-deception, but even if this is wrong, wishful thinking is a lot closer to the "self-deception" that Nietzsche attributes to the slaves than intentional action is.

The comment about the prudence displayed in the slaves' inversion of values actually works against the interpretation on which the slaves invert values by means of an intentional action. This "prudence of the lowest order" is something Nietzsche attributes to insects when they play dead in times of danger (GM I:13). Unless we are willing to say that Nietzsche sees insects as having the deliberative complexity that would make an attribution of prudence appropriate, we should regard this suggestion of intentional action as merely figurative. Indeed, Nietzsche often describes the non-intentional behaviors of organisms in ways that seem to attribute an inappropriately robust intentional background to them. The following aphorism, from *Twilight of the Idols*, is a nice example: "When stepped on, a worm doubles up. That is clever. In that way he lessens the probability of being stepped on again. In the language of morality: *humility*" ("Maxims and Arrows" 31). The talk of action and cunning in reversing values should, I think, be read as figuratively as talk of the clever humility of worms.

Rudiger Bittner (in his "Ressentiment") has offered an interpretation according to which the slaves' creation of values is part of a phenomenon of "sour grapes". Like the fox in Aesop's fable, who avoided being frustrated by his inability to get some ripe grapes by believing that they were actually sour, the slaves avoid being frustrated about their inability to enjoy the nobles' pleasures by believing that the nobles are mired in immorality for enjoying those pleasures. In this way, they reduce the frustration involved in believing that they are missing out on something good. Even more clearly than the account where the slaves invert values through conscious intentional action, this account will be congenial to my broader argument. It does not invoke moral properties to do any explanatory work, and it is clear that the "sour grapes" way is not a reliable process of belief-formation. I disagree with it as an interpretation of the first essay, however, because it does not put enough emphasis on the vengefulness of the slaves. It seems that Nietzsche wants to emphasize how the slaves make their moral values out of "revenge and hate" and out of the "hope for revenge" (GM I:14). While Bittner can adequately explain why Tertullian would predict a bad afterlife for the nobles, Tertullian's vehement hatred of the nobles and the way he exults in their suffering when he expresses his vision of the afterlife is better accounted for by a "vengeful thinking" interpretation than a "sour grapes" one. Bittner's interpretation is better suited to accounts of irrational belief formation about the afterlife that appear

elsewhere in Nietzsche's work, where vengefulness seems to play a less significant role.[4]

FN:4

Why does vengeful thinking bring the slaves to these particular beliefs about the afterlife and morality, but not to beliefs about other issues more directly related to the happiness of the nobles? One reason for the special potency of vengeful thinking in the cases I have described has to do with the difficulty of disconfirming the kinds of beliefs that the slaves do form. The slaves cannot easily form the belief that the nobles are starving to death, even though vengeful thinking might dispose them to embrace such a belief. That belief would be empirically disconfirmed when the slaves saw the nobles feasting and enjoying themselves, and when they saw that the nobles were not gaunt and emaciated. The situation is different with beliefs about the afterlife and beliefs about general principles of morality. These beliefs, if not totally isolated from empirical confirmation and disconfirmation, are at least difficult to confirm or disconfirm empirically. Evidence does not get in the way of vengeful thinking on these topics, while evidence can impede the belief-formation that arises from vengeful thinking about other issues.

All Nietzsche offers in the First Essay is an account of the historical origins of contemporary Western moral values. He does not present an explicit account of how this moral system was transmitted from one generation to the next, so that it could become the dominant moral system of our time. It is clear, though, that he thinks such transmission took place, and that his account needs to be supplemented by some story about how moral norms are handed down through the centuries so that we can inherit slave morality. An interpretation according to which *ressentiment* has generated new moral systems so frequently as to eliminate the need for mechanisms transmitting the slave morality from Roman times to us would not need such supplementation, but the text does not support it. Nietzsche does not present his account in the First Essay as one instance of a process that occurs again every time people come to believe in slave morality. Certainly, Nietzsche does hold that *ressentiment* has affected people's values at other points in history—the French Revolution (GM I:16), anarchism, and anti-Semitism (GM II:11) provide examples. But he does not give an account of a contemporary slave revolt that generated our entire system of values, and he does not suggest that our entire system of values is the result of one. Instead, he claims that the victorious values of today arose in the revaluation of two millennia ago, which he spends the majority of the First Essay describing—"Israel, with its revenge and revaluation of values, has thus far again and again triumphed over all other ideals, all more *noble ideals*" (GM I:8). When Nietzsche's opponent, the modern democrat, says that the slave revolt "happened through the Jews . . . never has a people had a more

[4] Examples include the formation of beliefs about heaven discussed in *Zarathustra* I: "On the Afterworldly" and in the third essay of the *Genealogy*.

world-historic mission," Nietzsche seems to regard this as a correct characterization of his view (GM I:9). The Jewish revaluation would not have been so world-historic if it was just one among many revaluations.

How was the slave morality passed down through the generations so that we moderns could inherit it? While Nietzsche does not provide an account of this process, one can be developed for him. One just needs to deploy whatever account one has of the processes by which parents transmit their moral values to their children, and by which societies cause their members to internalize prevailing moral values. The institutional power of the Church to perpetuate Christian morality is also a significant factor. And given the difficulties involved in finding genuine disconfirming evidence for one's beliefs about morality, one can see why the passage of time did not result in the accumulation of sufficient counterevidence to overturn the moral system that arose when the slaves completed their revaluation. Even if counterevidence had been available, it might not have been given sufficient attention under the intellectual stagnation that dominated most of the centuries between Tertullian and ourselves. It took well over a thousand years before Galileo discovered the simple physical fact that falling objects of different weights accelerate at a uniform rate, and moral-theory-disconfirming observable facts would probably have had a similarly difficult time attracting notice.

Now that I have presented Nietzsche's account of the origins of morality, I will turn my focus to the arguments against non-reductive naturalistic moral realism that it supports. Nietzsche's account supports the Explanatory Argument by furthering explanations of our moral observations that do not invoke any moral properties. His account supports the Unreliability Argument by making us aware that moral beliefs arise through a process that does not reliably generate true belief. After giving some background on the debates that these arguments will enter into, I will describe the arguments themselves in detail.

One way to defend moral realism is to argue that moral properties have a place in the best explanatory picture of the world. The explanatory power of moral properties is attacked by anti-realists like Harman and Leiter, and defended by naturalistic moral realists like Nicholas Sturgeon and Geoffrey Sayre-McCord. When these naturalistic moral realists discuss the existence of moral properties, they are talking about irreducible moral properties.[5] In "Moral Explanations," Sturgeon argues that it is

FN:5

[5] The reality of reducible moral properties is a less complicated issue—one just has to establish the reality of the reduction base, and reductive moral realists usually offer reduction bases whose existence is uncontroversial. The challenge for reductive moral realists is finding a reduction base that fits the semantics of moral discourse. In writing about "moral properties" here, I will generally be referring to irreducible moral properties.

"plausible to cite moral facts as part of an explanation of non-moral facts, and in particular of people's forming the moral opinions that they do" (243). In saying this, he responds to Harman's view (in "Ethics and Observation") that in cases where we make the observation that some act was morally wrong, "neither the moral principle nor the wrongness of the act can help explain why you observe what you observe" (123). If what justifies us in believing in some property is that the property figures in the best explanatory picture of the world, Sturgeon's view will have us being justified in believing in moral properties while Harman's view will not.

In "Moral Theory and Explanatory Impotence," Sayre-McCord lays out several versions of the "Explanatory Criterion" which is being used in determining whether or not we are justified in believing that moral properties exist. The version of the Explanatory Criterion that he eventually relies on in his discussion of the issue is the following: "A hypothesis should not be believed if the hypothesis plays no role in the best explanation we have of our making the observations we do" (267).

Sayre-McCord attacks the anti-realists' Explanatory Argument, which can be expressed as follows:

(P1) The hypothesis that moral properties exist is not part of the best explanation of our observations.
(P2) If a hypothesis is not part of the best explanation of our observations, we are not justified in believing it.
(C) We are not justified in believing that moral properties exist.

P2 is a rephrasing of Sayre-McCord's Explanatory Criterion, accepted by both the realists and the anti-realists in this part of the debate. The disagreement between realists and anti-realists is about P1. Harman and Leiter believe that the hypothesis that moral properties exist does not figure in the best explanations of our observations; Sturgeon and Sayre-McCord believe that the hypothesis has an important role to play.

Consider a set of observations which moral realists might claim their hypothesis usefully explains—observations about the moral rightness and wrongness of various actions. If moral observations are construed in a parallel way to our observations about color, the case for moral realism might begin to look as strong as the case for realism about colors.[6] To many, it seems plausible to give a realist explanation for our color experiences on which my experience of red is explained by the property of redness being instantiated in a strawberry that is in front of me. Similarly, it can seem

FN:6

[6] While realism about objective color properties is a popular position, it is far from uncontroversial. Furthermore, many realist positions—for example, the reductive physicalist position of Frank Jackson—won't supply the non-reductive moral realist with a comfortable analogy to the position he wants for moral properties.

plausible to give a realist explanation for our moral observations. If I were asked what I thought of Churchill's decision to block food imports into 1940s Bengal and allow millions of Indians to starve, I would respond that it was viciously wrong. In giving a realist explanation, one would assert that Churchill's actions had the property of being viciously wrong, and that this explained the belief I formed in thinking about the situation. (Here I do not fill out the realist explanation—most likely, there would be steps between Churchill's action and my judgment. I am just presenting the essential features of a realist explanation.)

Harman defends P1 of the Explanatory Argument, saying that moral properties are not necessary to explain moral observations. All we need for a good explanation are the non-moral facts—namely, that Churchill caused millions to starve, that I have learned the history of India, and that I have been acculturated to think that causing mass starvation is wrong. These factors are sufficient to explain my making the observations I do. As Harman writes, "an assumption about moral facts would seem to be totally irrelevant to the explanation of your making the judgment you make" (122).

FN:7

Moral judgments are not the only things that moral properties are said to explain[7]. Sayre-McCord writes that

> Many very useful, and frequently offered, explanations of events in the world (and so of our observations of those events) make reference to moral facts. Mother Teresa's goodness won her a Nobel Prize; Solidarity is popular because of Poland's oppressive political institutions; millions died in Russia as a result of Stalin's inhumanity; people are starving unnecessarily because of the selfishness of others; unrest in Soweto is a response to the injustice of apartheid. (275)

However, explanations of these events that appeal only to psychological, social, and physical properties seem to have satisfied scholars who study them. As Brian Leiter points out, "moral facts appear to play no role in any developed explanatory theory . . . there is no school of 'Moral Historians' using moral facts to do any interesting or complex explanatory work" (94). Here Leiter draws on existing explanatory accounts of historical and social phenomena to argue for the first premise of the Explanatory Argument.

[7] Even if it turned out that properties like honesty played a role in the best explanatory theories, that would not spell victory for the non-reductive moral realist. A reductive moral realist who wanted to reduce moral goodness to honesty would be able to declare victory, because the reduction base for morality had earned its explanatory credentials. (As few philosophers since Diogenes have been eliminativists about honesty, defending its reality by pointing out its role in best explanations might be unnecessary in contemporary debates.) But a non-reductive moral realist would have to show something further—that the moral value supervening on honesty played an important explanatory role. If aspects of honesty other than its moral value did all of the explanatory work, the existence of the moral value would not be successfully vindicated. I thank Geoffrey Sayre-McCord for helping me see this.

Nietzsche endorses the Explanatory Argument in an early passage of *Daybreak*. He claims that giving non-moral causal explanations of moral phenomena will eliminate the need to posit moral properties:

> *Sense for morality and sense for causality in counteraction*—In the same measure as the sense for causality increases, the extent of the domain of morality decreases: for each time one has understood the necessary effects and has learned how to segregate them from all the accidental effects and incidental consequences (*post hoc*), one has destroyed a countless number of *imaginary causalities* hitherto believed in as the foundations of customs—the real world is much smaller than the imaginary . . . (10)

In this passage, "sense for causality" refers to the ability to identify correctly the causal processes explaining events. When Harman gives non-moral explanations of moral observations and Leiter cites non-moral explanations of historical phenomena, they work to destroy the "imaginary causalities" that Sturgeon and Sayre-McCord defend. With the moral properties eliminated, our overall explanatory picture of the world comes out to be "smaller." In a later work, Nietzsche describes our lack of justification for holding moral beliefs once the explanatory irrelevance of moral properties has been demonstrated as "the moral problem: *Why have morality at all* when life, nature, and history are 'not moral'?" (*Gay Science* 344).[8]

FN:8

Nietzsche's explanation of how the slaves arrived at their moral beliefs in the *Genealogy* is one that moral anti-realists who accept the Explanatory Criterion could be happy with. It provides the support needed for the first premise of the Explanatory Argument. At no point does one need to invoke a moral property to explain anything. In order to explain how the slaves got their beliefs in the immorality of aggressive and violent actions like the ones the nobles often performed, we do not need to assume that an irreducible property of being immoral supervenes on the nobles' aggressive and violent behavior. (Nietzsche's unwillingness to attribute the property of immorality to the nobles' violent and aggressive behavior is one reason why the First Essay can be so unsettling to read.) All we need to assume is that the nobles did things that caused the slaves to desire vengeance, and that vengeful thinking caused the creation of slave morality in the way that I have described. To generate an explanation of present-day moral beliefs, we likewise need no reference to irreducible moral properties. We just need to discover the psychological properties that are involved in the acquisition of moral beliefs from other members of society.

This account of how moral beliefs originated will help the anti-realist explain moral observations by explaining how we acquired the "auxiliary hypotheses" that give these observations their moral content. Consider the way scientific observations

[8] In this section, Nietzsche suggests a parallel between the metaphysical status of epistemic and moral norms that is very similar to the parallel Sayre-McCord suggests later in his essay.

go—scientists encounter situations where some real entity (say, a proton) causes observable phenomena and apply their beliefs about the kind of entity that would be causally responsible for the observed phenomena, resulting in the formation of beliefs about the real entity. If the scientist's auxiliary hypotheses connecting the observed phenomena to the real entity are good, entities of the same kind as the real entity will probably have played a role in the construction of the auxiliary hypotheses. Harman's original example involves a scientist who sees activity in a cloud chamber and infers that a proton must have moved through the chamber. If the scientist is correct in holding the auxiliary hypothesis that such activity in a cloud chamber is caused by protons, the activity of protons probably plays some role in explaining why he accepts this hypothesis. Most likely, previous experiments establishing a link between protons and similar motions in cloud chambers played a role in causing his acceptance of the auxiliary hypothesis. So protons play a key role in explaining why the scientist holds the physical theory he does, in addition to explaining the events in the cloud chamber.

Now consider the way that a Harman-style explanation of moral observations would go—observers encounter situations constituted entirely by non-moral facts and apply their beliefs that certain moral properties supervene on the non-moral facts, resulting in moral observations. A realist might ask for an explanation of how the beliefs connecting moral properties to non-moral facts—the analogs of auxiliary hypotheses in the scientific case—arose. Nietzsche offers an explanation that establishes a disanalogy between the moral case and the scientific case. While protons need to be invoked to explain the existence of auxiliary hypotheses according to which protons are connected to events in cloud chambers, no moral properties need to be invoked in explaining why we hold auxiliary hypotheses according to which moral properties are connected to natural properties. To do the explanatory work, we just need violent aggression directed by nobles towards slaves, some vengeful thinking by the slaves, and the means to transmit the results of vengeful thinking down through the ages.

Evaluating the correctness of Nietzsche's historical account is beyond the scope of this essay. Of course, his particular account need not be correct for the Explanatory Argument to go forward. Nietzsche's account supports the Explanatory Argument by helping to explain our moral observations—in particular, accounting for the existence of the auxiliary hypotheses that they depend on—without invoking moral properties. Nietzsche's story is by no means the only one that could do this work. For example, if some Marxist account is correct in describing how economic relations give rise to the prevalent moral systems from which our auxiliary hypotheses are drawn, it might underwrite the first premise of the Explanatory Argument just as well.

Now I will move towards a discussion of the Unreliability Argument. The Unreliability argument comes into play if P1 of the Explanatory Argument is accepted, but P2—the Explanatory Criterion—is rejected. We might reject P2 if realists could show that we were justified in believing hypotheses according to which some properties

existed, but did not play a role in our best explanations. In that case, anti-realists would have to offer a new argument which would rule out moral properties but not the properties cited in the realist argument that forced us to reject P2. This is what the Unreliability Argument will do.

While Sayre-McCord wants to contest the anti-realists' claims by arguing that moral facts cannot be dropped from the best explanations, he suggests a fallback position to defend if satisfactory non-moral explanations are available. Immediately after the passage where he claims that geopolitical events can be explained with reference to moral properties, he continues, "Even if such explanations could be replaced by others that appeal only to psychological, social, and physical properties, without mention of moral facts, the moral explanations would still be useful in just the way talk of colors remains useful even in light of theories of light" (275). Even if reductions of color properties fail, it is not clear that non-reductive realism will triumph over anti-realism with regard to color. One might hold fast to the Explanatory Criterion and argue that we really are unjustified in positing color properties. Perhaps we should be error theorists about color, and give a similarly anti-realist account of moral talk[9].

FN:9

At the end of his essay, Sayre-McCord presents another set of candidate irreducible properties that can be analogized to moral properties, just as color properties can. These are the properties we appeal to in evaluating explanations. As Sayre-McCord argues, one cannot use the Explanatory Criterion against moral properties unless one has already accepted epistemic value properties in virtue of which explanations excluding moral properties are better than explanations including them: "If one explanation is to be better than another in virtue of being simpler, more general, more elegant, and so on, then simplicity, generality, and elegance cannot themselves be evaluatively neutral. Were these properties evaluatively neutral, they could not account for one explanation being better than another" (277). An error theorist about epistemic value who still wants to use the Explanatory Argument will not be able to explain why we are justified in accepting anti-realist explanations over realist ones. At least on this count, epistemic value is a better "partner in crime" for moral properties than color is.[10]

FN:10

Depending on how the explanations go, there may be grounds for an objection to Sayre-McCord's fallback position. Suppose we discover, as we trace the genealogy of our moral beliefs, that they arose through an unreliable process of belief formation—a process that resulted in the formation of a number of other false beliefs. Then we would

[9] It is hard to see how the color–morality analogy would provide a powerful argument for non-reductive realism, since neither reductive realism nor anti-realism about color can be dismissed out of hand.

[10] That is, assuming that reductive realism does not work. Also, while error theory is not an available option here, one might still try to defend a non-cognitivist theory of epistemic value against Sayre-McCord, and argue for epistemic norms without the ontological commitments of realism.

have an additional reason to withhold belief in moral properties. Beliefs resulting from reliable processes are likely to be true, while beliefs resulting from unreliable processes are not. If we did not have unjustifying explanations of how we attained our color-beliefs and our epistemic value-beliefs, casting the reliability of these processes of belief-formation into doubt, belief in color properties and epistemic value properties would be justified.

This brings us to the Unreliability Argument against moral properties:

(P1) Our beliefs in moral properties were caused by a historical process that is unreliable in generating true belief.
(P2) We know whether an unreliable historical process caused our beliefs in moral properties.
(P3) Beliefs that we know to be caused by unreliable historical processes are unjustified.
(C) We are not justified in believing that moral properties exist.

If we get an independent source of justification for our moral beliefs—as we might if we discovered that moral properties played a role in the best explanations, defeating the Explanatory Argument—the Unreliability Argument will not obtain its conclusion. If the Explanatory Argument is false because moral properties actually do have explanatory power, and if our moral beliefs were caused by our seeing that moral properties have explanatory power, P1 of the Unreliability Argument will be false. (I intend 'cause' in P1 to be read so that beliefs that were initially caused by an unreliable process but which were then reinforced by a reliable process will count as reliably caused. In cases where one's doxastic state is causally overdetermined, one reliable cause will make the belief reliably caused.) Believing hypotheses because they are part of the best explanation of our observations is a reliable way of forming beliefs, and if we discover that moral properties figure in the best explanations, the causal influence that this discovery exerts by backing up our belief in moral properties will be sufficient to justify us in believing that they exist.

Even for a reliabilist, the mere existence of an unreliable step way back in the causal history of the belief may not be enough to make the belief unjustified, if we're unaware of the step and it's sufficiently disconnected from our acquisition of the belief. If we learned our moral beliefs from our parents, and most of the things our parents tell us are true, accepting our parents' beliefs may count as a reliable process. In this case, moral beliefs may be reliably formed regardless of the irrationality of their historical origin. (While it's unclear what the best way to individuate kinds of belief-forming processes is, reliabilists will generally put more emphasis on proximal factors than distal ones.) The view that the majority of contemporary people are currently justified in holding moral beliefs, because of the reliability of proximal belief-forming processes and despite the unreliability of the historical processes, is compatible with the soundness of the Unreliability Argument if the argument is

applied only to the beliefs of a historically knowledgeable minority. If a reliable source actually makes everyone aware that their moral beliefs were formed in an unreliable way, even because of a step way back in history, this will eliminate the justification for continuing to hold them. One only loses one's justification for believing in irreducible moral properties when one becomes justified in believing an account on which belief in these properties is unreliably caused. A Nietzschean story about the origins of morality may not point out a pre-existing lack of justification for believing in moral properties—it can actually take one's justification away. This is why P2 is necessary to the Unreliability Argument.[11]

FN:11

The appeal of P3 should not be limited to those who accept a version of reliabilist epistemology. Coherentists will accept P3 because maintaining belief in moral properties will cohere badly with the belief that the process that generated them is unreliable. Foundationalists will accept P3 because unreliable processes are unlikely to provide either foundational or inferential justification for beliefs. Those who accept a Nozick-style tracking account of justification will accept P3 as long as the process of belief-formation is unreliable enough that our beliefs would not track the existence of moral properties in counterfactual situations. Since the process that generated the slaves' moral beliefs would have brought about such beliefs even without the existence of the moral properties that would have made them true, these beliefs will be unjustified according to Nozick's theory.

Probably the best general way to understand P3 is as an instance of defeasible reasoning—specifically, as a case in which we have found an undercutting defeater for one of our beliefs. Undercutting defeaters give us reason to reject the conditionals tying putative evidence to the propositions that they are putative evidence for:

If P is a *prima facie* reason for S to believe Q, then R is an undercutting defeater for P iff R is a reason for S to believe that it is not true that P would not be true unless Q were true.[12]

FN:12

We might take the fact that we make moral observations as *prima facie* evidence for the existence of moral properties. But if we see that the historical process that generated our moral observations does not reliably indicate that these properties exist, we will have an undercutting defeater for the existence of the moral properties. When we discover the absence of a reliable connection between the observations and the properties, we will be justified in rejecting the belief that the observations would not be made unless the properties existed.

If we come to know that Nietzsche's account from the First Essay of the *Genealogy* is correct, P1 and P2 of the Unreliability Argument will be true. P1 will be true because

[11] I thank Anthony Gillies for assistance on this point.

[12] Scott Sturgeon 102, adapting the terminology from John Pollack's *Contemporary Theories of Knowledge*.

the process by which we came to our beliefs—inheriting the outcome of past vengeful thinking—is clearly unreliable. Wishful thinking is generally unreliable, and we have an example of the unreliability of the slaves' particular brand of wishful thinking, in their absurd beliefs about the nature of the afterlife. Inheriting beliefs formed in this way is not a reliable way of getting at the truth. The fact of our knowing this account will make P2 true. Then if the epistemic norms we accept lead us to P3—as most contemporary theories of epistemic justification do—the Unreliability Argument will obtain its conclusion. We will be unjustified in believing in moral properties.

Now I will examine the implications of the Unreliability Argument for color and epistemic value properties. If the reduction of color properties fails, will colors be ruled out of our ontology along with morality because of the Unreliability Argument? There is no reason to think that they will. If we replace "moral properties" with "color properties," P1 becomes false. Beliefs caused by sense-perception under normal conditions are, in general, true. Most of our color perceptions fall into this category, and therefore it makes sense to believe in colors. That is something that cannot be said of beliefs caused by vengeful thinking, or wishful thinking in general. Vengeful thinking is not a reliable cause of true belief. So any concern that the rigorous requirements of the Explanatory Argument will force us to anti-realist conclusions about color properties is satisfactorily dealt with by the softer requirements of the Unreliability Argument.

Epistemic value properties will most likely pass the tests offered by the Unreliability Argument. Even if reductive accounts of epistemic value fail, many approaches to epistemology come with plausible stories about how the theories of justification that they issue are reliably caused. Consider, for example, naturalistic epistemologists who develop their views about epistemic value by studying sciences that yield correct predictions. They would be able to point to their study of science in defending the processes by which they arrived at their theories of epistemic value. At the very least, the Unreliability Argument will not support a global attack on any theory of epistemic justification[13]. If the skeptic thinks we are justified in accepting the conclusions of the Unreliability Argument with respect to epistemic value properties, this positive claim about justification—which presupposes epistemic value properties that can underwrite our justification for accepting the skeptical view—will contradict the global anti-realism, much as Sayre-McCord suggested.

FN:13

It is possible—though unlikely—for all the premises of both the Unreliability Argument and the Explanatory Argument to hold in a world where moral properties

[13] While the Unreliability Argument doesn't support global error theory about epistemic value, it might be locally effective against some views, given an appropriate historical account. Suppose we discover that Alvin Goldman has been using subliminal messages to make us unthinkingly profess reliabilist intuitions when presented with his examples. Then we might be justified in reducing our credence in reliabilism, at least until new examples could be generated.

actually exist. (While the truth ratio of wishful thinking would rise in such a world, it might not rise far enough to falsify P1 of the Unreliability Argument and make wishful thinking count as a reliable process.) In this case, we would have unjustified true belief in the existence of moral properties, and so fall short of moral knowledge. This should not be counterintuitive. If wishful thinking causes me to believe that I will make a fortune on my trip to Las Vegas, and amazing luck at the craps table brings me great wealth, my prior belief will still be unjustified, even though it is true. To put the issue in terms of Nozick's tracking account of epistemic justification, vengeful thinking would still have us believing in moral properties, even in the worlds where the moral properties did not exist. So despite the truth of our beliefs, we are not justified in believing in moral properties.

As with the Explanatory Argument, Nietzsche's particular account need not be true in order for the Unreliability Argument to go forward. All that is needed is some story according to which moral beliefs arise as the result of an unreliable historical process. If our moral beliefs are the result of attempts by powerful individuals to encourage the beliefs that would help them consolidate their power, and if these attempts are generally unreliable sources of true belief, P1 of the Unreliability Argument will be vindicated.

If the two arguments of this essay are successful, and if Nietzsche's account advances them in the way I have suggested it can, we can see how a genealogy of morality bears on the justificatory status of our moral beliefs. To assert that the causal origins of our moral beliefs are relevant to their justificatory status is not necessarily to commit the Genetic Fallacy. If the things that make a belief true play a role in the best explanation of its genesis, we are justified in continuing to hold the belief. If they play no role in the best explanation of its genesis, a potential source of justification is eliminated. And if we find that the causal processes that historically gave rise to the formation of the belief are unreliable, this may destroy our justification for continuing to believe.

The combination of the Explanatory Argument and the Unreliability Argument set out in this essay is not necessarily doom to the non-reductive moral realist. For one thing, maybe Nietzsche's account of the origin of morality is wrong, and vengeful thinking did not play such an important role in the formation of moral belief. That would allow realists to set aside the Unreliability Argument and eliminate one possible explanation for how our moral beliefs were formed. It might also turn out that unreliable processes are implicated in the formation of some of our moral beliefs, while reliable processes are implicated in others. This would have interesting ramifications for the methodology of normative ethics. Rather than just building theories to fit the intuitions generated by various cases, moral theorists would have to ask whether the intuitions arose from reliable causes or unreliable ones. If a theory failed to fit prevalent moral intuitions that arose from unreliable causal processes,

this would not be a flaw. But if there is a plausible explanation of all our moral beliefs in terms of vengeful thinking or some other unreliable process that does not involve moral properties, non-reductive moral realists have reason to worry. If such an explanation is in fact the best one, and if it is presented to them, they face the full force of both anti-realist arguments covered in this essay.

References

Bittner, Rudiger, "Ressentiment", in Richard Schacht (ed.), *Nietzsche, Genealogy, Morality* (Berkeley: University of California Press, 1994).

Harman, Gilbert, "Ethics and Observation", in Geoffrey Sayre-McCord (ed.), *Essays on Moral Realism* (Ithaca: Cornell University Press, 1988).

Leiter, Brian, "Moral Facts and Best Explanations" (Social Philosophy and Policy Foundation, 2001).

Nietzsche, Friedrich, *Beyond Good and Evil*, trans. Walter Kaufmann (New York: Random House, 1966).

——— *Daybreak*, trans. R. J. Hollingdale (Cambridge: Cambridge University Press, 1997).

——— *The Gay Science*, trans. Walter Kaufmann (New York: Vintage, 1974).

——— *On the Genealogy of Morality*, trans. Maudemarie Clark and Alan Swensen (Indianapolis: Hackett, 1998).

——— *Thus Spoke Zarathustra*, trans. Walter Kaufmann (New York: Penguin, 1982).

Sayre-McCord, Geoffrey, "Moral Theory and Explanatory Impotence", in Geoffrey Sayre-McCord (ed.), *Essays on Moral Realism* (Ithaca: Cornell University Press, 1988).

Sturgeon, Nicholas, "Moral Explanations", in Geoffrey Sayre-McCord (ed.), *Essays on Moral Realism* (Ithaca: Cornell University Press, 1988).

Sturgeon, Scott, "Truth in Epistemology", *Philosophy and Phenomenological Research* 51/1 (1991): 99–108.