5-7-2011

Nietzsche's Causally Efficacious Account of Consciousness

Bradley Wissmueller
Georgia State University

NIETZSCHE'S CAUSALLY EFFICACIOUS ACCOUNT OF CONSCIOUSNESS

by

BRADLEY WISSMUELLER

Under the Direction of Dr. Jessica N. Berry

ABSTRACT

Many interpreters read Nietzsche as an epiphenomenalist. This means that, contrary to everyday "felt" experience, consciousness has no causal influence on our actions. In the first half of this paper I show that an epiphenomenalist interpretation proposed by Brian Leiter is unsupported by Nietzsche's texts. Further, contemporary research does not conclusively support epiphenomenalism, as Leiter claims. In the second half of the paper I present the novel, causally efficacious view of consciousness that is supported by Nietzsche's texts. This view of consciousness does not present consciousness as a self-caused faculty that is in some way separate from the rest of our mind and body, but rather views consciousness as a non-essential property of certain mental states. I trace the development of this idea through two key passages and show that, in the danger it presents as well as in the promise, consciousness is clearly causally efficacious.

INDEX WORDS: Consciousness, Nietzsche, Epiphenomenalism, Brian Leiter, Paul Katsafanas, Daniel Wegner, Benjamin Libet, Edward Nahmias, Alfred Mele.

NIETZSCHE'S CAUSALLY EFFICACIOUS ACCOUNT OF CONSCIOUSNESS

by

BRADLEY WISSMUELLER

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Arts

in the College of Arts and Sciences

Georgia State University

2011

NIETZSCHE'S CAUSALLY EFFICACIOUS ACCOUNT OF CONSCIOUSNESS

by

BRADLEY WISSMUELLER

Committee Chair: Dr. Jessica N. Berry

Committee: Dr. Eddy Nahmias

Dr. Sebastian Rand

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

May 2011

ACKNOWLEDGEMENTS

First, I would like to thank Jessica N. Berry for spending many hours with me over the last year working through the ideas in this paper and several other issues regarding Nietzschean action. Throughout the process she has offered important criticisms of my developing interpretation of Nietzsche, kept me motivated and on track to complete this paper, and helped me improve the quality of my writing. I would also like to thank Walt Duhaime and Bryan Russell, who participated in a discussion group on Nietzschean action which also included Dr. Berry and myself; Eddy Nahmias and Kyle Walker, who provided helpful feedback on two earlier drafts of a paper which became the basis for this thesis; Sebastian Rand, who provided thoughtful comments on the penultimate draft; and Eleanor Johnson, who proofread two of the early drafts of this paper, kept me from getting too far away from "common sense," and helped me through several stressful days leading up to the various thesis deadlines.

TABLE OF CONTENTS

LIST OF FIGURES

I. INTRODUCTION

Epiphenomenalism is the view that our consciousness has no causal power. This means that consciousness can never influence our behavior causally, in spite of what we take to be first-hand experience of our conscious will directly affecting our behavior. George Graham, in his introduction to the philosophy of mind, stresses the radical nature of this view: "I find epiphenomenalism...so grossly implausible that I cannot imagine how anyone could embrace it" (1993: 185). He then quotes Terence Horgan, who echoes this sentiment: "Epiphenomenalism...should be an utter last resort, to be embraced only if all viable alternatives prove to be...paradoxical and untenable" (Graham 1993: 185). It is no small matter, then, whether Nietzsche holds this view. While it would come as no surprise that Nietzsche holds a radical view, if Nietzsche does indeed view consciousness as epiphenomenal, it would significantly inform how we should read and understand some of the main themes of his philosophy.[1]

While Nietzsche's specific views on consciousness have only recently been addressed in detail in the secondary literature, the conclusion that Nietzsche is an epiphenomenalist, in some sense of the term, has found support among several Nietzsche scholars. Brian Leiter can be seen as a representative of this reading, both in his 2002 book, *Nietzsche on Morality* and in his 2009 article, "Nietzsche's Theory of the Will." One of the primary motivations for his reading, presented in the book, is Nietzsche's naturalism (Leiter 2002: 2-8). As Leiter points out, Nietzsche is concerned with showing that we are not of a "'higher...[or] of a different origin' from the rest of nature" (Leiter 2002: 7). Nietzsche gives several explanations of human action that (at least initially) seem not to require a causally efficacious consciousness at all. Many of Nietzsche's references to consciousness attribute to consciousness some sort of *causa sui* ability, which Nietzsche clearly rejects as absurd. There is not some special human consciousness that is re-

[1] Or, even, possibly, whether we should read or try to understand his texts at all!

sponsible for what we do; rather there is a psycho-physiological explanation that involves the same "stuff" that animals and even plants have. According to Leiter, when we consider ourselves as different from nature because we believe ourselves to possess some self-causing conscious faculty, Nietzsche thinks we are fooling ourselves: there is no faculty standing behind all action and thought, the idea of consciousness is just an epiphenomenon.

My paper has two main sections. In the first section, I will look at Leiter's view as it has most recently been presented. I discuss both the textual evidence Leiter cites and the contemporary empirical research Leiter uses in support of the claim that we should charitably interpret Nietzsche as an epiphenomenalist. After showing that this epiphenomenalist interpretation is inconclusive at best, I will turn to the second section, which will make up the bulk of my paper. In this second section, I will argue that there is a coherent way to make sense of Nietzsche's disparate references to consciousness. This reading begins with a fundamental distinction between texts in which Nietzsche discusses consciousness as a self-caused faculty and those in which Nietzsche discusses consciousness as a non-essential property of certain mental states. By supporting what Leiter says about the former, we can still read Nietzsche as a naturalist who views our development as continuous with the rest of nature. In fact, in order to understand Nietzsche's view of consciousness as a non-essential property of certain mental states, I will turn to two key passages in Nietzsche, *The Genealogy of Morals* II 16 and *The Gay Science* 354, which show how a non-epiphenomenal, causally efficacious consciousness developed from man's "unconscious animal past." On the view I present, consciousness is not a special, self-caused faculty, but it still points to an important mental difference between human animals and others: humans have the ability to conceptualize. After explaining what this means and defend-

ing this interpretation, I will close by making brief remarks about the way in which this view can inform how we read and understand Nietzsche's central ideas.

II. LEITER'S EPIPHENOMENALIST INTERPRETATION

In his recent work, Brian Leiter has argued for the view that Nietzsche is an epiphenomenalist. His 2009 article, "Nietzsche's Theory of the Will," argues specifically for a reading of Nietzsche as a will-epiphenomenalist who holds the view that there is "no causal link between the experience of willing and the resulting action" (Leiter 2009: 123). More specifically, Leiter equates this with token-epiphenomenalism, which he defines in his book, *Nietzsche on Morality*, as the view that "conscious states are simply *effects* of underlying type-facts about the person, and play no causal role whatsoever" (2002: 92). Leiter defines type-facts as "either *physiological* facts about the person, or facts about the person's unconscious drives or affects" (91). There are three main arguments that Leiter derives from Nietzsche's texts to support this interpretation: (1) an argument from the phenomenology of willing, (2) an argument from false causality, and (3) a positive argument about "the real genesis of action" (2009: 121). I will consider each of these arguments individually to determine whether they do in fact provide evidence for epiphenomenalism. Even with these arguments, however, Leiter admits that Nietzsche is "generally ambiguous as to which view of the will he decisively embraces": will-epiphenomenalism, or a view in which the will is a secondary cause. By this latter view, Leiter means that while one's conscious will is the immediate cause of the action, one's type facts are the *primary* cause of the action since they cause the conscious willing (see Figure 1.1 below).

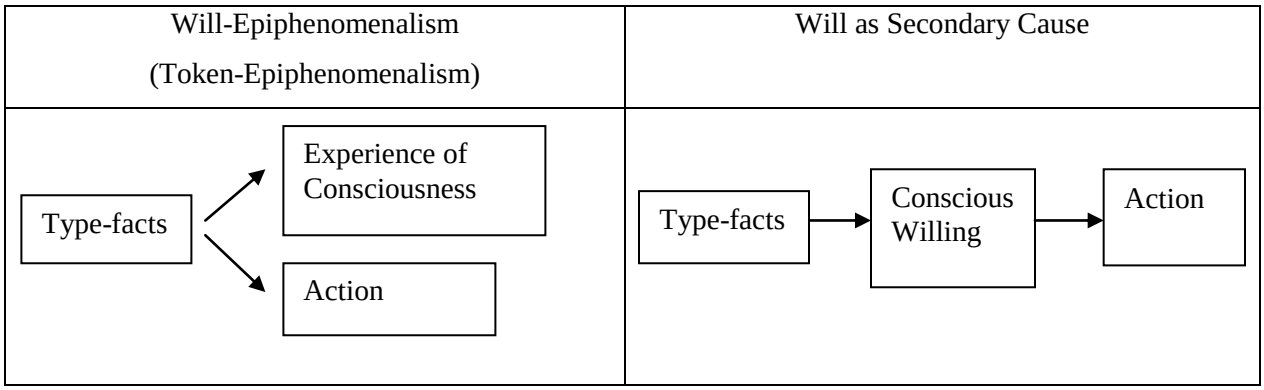

Figure 1.1 The Two Options Leiter Presents

To decide between these two views, Leiter presents the recent empirical research of Daniel Wegner and Benjamin Libet, which appears to support epiphenomenalism, and argues that Nietzsche is most charitably read as a token-epiphenomenalist. After looking at Leiter's three main arguments and the textual evidence he uses to support them, I will discuss the empirical research that Leiter cites in support of his interpretation. My conclusion in this section will be both that (1) the textual arguments Leiter cites, are, at the very best, more ambiguous than he lets on, and, at worst, they support a view contrary to Leiter's reading, and that (2) the interpretation of the empirical data Leiter uses to support token-epiphenomenalism is too controversial to support an argument from interpretive charity.

1. The Textual Arguments

A. *Beyond Good and Evil 19*

Section 19 of *Beyond Good and Evil* has been interpreted in a wide variety of ways. Leiter reads the entire passage as a presentation of a phenomenology of willing. As Leiter notes, Nietzsche says there are three main experiences involved in what we uniformly refer to as "the will." The first is a "plurality of sensations" (the sensation of moving away from something, to-

wards something else, and the sensation of muscular motion); second is thinking or "a ruling thought"; and third is "the affect of the command" (BGE 19).[2] Referencing an argument against the Cartesian "I" from two sections earlier, Leiter claims Nietzsche attacks the second experience mentioned—the ruling thought. In that section, Nietzsche points out that "a thought comes when 'it' wishes, and not when 'I' wish" (BGE 17). It is a "falsification of the facts," Nietzsche continues, "to say that the subject 'I' is the condition of the predicate 'think.' *It* thinks; but that this 'it' is precisely the famous old 'ego' is, to put it mildly, only a supposition, an assertion and assuredly not an immediate certainty" (BGE 17). Leiter attempts to draw from this section that "if we don't experience our thoughts as willed, then it follows that the actions that follow upon our experience of willing (which includes those thoughts) are not caused in a way sufficient to underwrite ascriptions of moral responsibility" (Leiter 2009: 113). Leiter here acknowledges that this conclusion actually relies on two premises. The first premise—that the ruling thought is causally determined by something other than the conscious will, since all of our thoughts are causally determined by something other than our consciously willing them into existence—is the one he thinks is demonstrated in *Beyond Good and Evil* 17. The second premise is that "being self-caused is a necessary condition for responsibility" (Leiter 2009: 113). The first premise is the focus of this paper and what is relevant to Leiter's main thesis—that Nietzsche is a token-epiphenomenalist. This first premise, however, does nothing to distinguish between a token-epiphenomenal view of the will and a view according to which the will is *secondarily* causal, since neither of these views posits the will as "self-caused."[3]

[2] I follow Walter Kaufmann's translation unless otherwise noted.

[3] The reason Leiter discusses moral responsibility and adds premise two is that he believes "it is only certain philosophers who think the need to be a self-caused agent is superfluous, something that can be finessed via some adroit dialectical moves" (Leiter 2009: 115). While it may be Nietzsche's view (and Leiter's understanding of the free will debate) that being self-caused (*causa sui*) is important for moral responsibility, there are, in fact, many compatibil-

So, even if our conscious will does not control what thoughts come into our mind, it is possible that once those thoughts are in our mind, the conscious will can affect them in a causally important way. Leiter's reading of this passage, which is compatible with both of the interpretations he gives, is by no means uncontroversial, and many other Nietzsche interpreters have given drastically different accounts of what Nietzsche is up to in this passage. For instance, Maudemarie Clark and David Dudrick read Nietzsche as offering a positive account of willing in this same passage (Clark and Dudrick 2009: 247-268).

B. *False Causality*

The second textual argument that Leiter presents is based on what Nietzsche says about "false causality" in *Twilight of the Idols* (Leiter 2009: 121). Here, Leiter points to a section called "The Four Great Errors." In this section, Nietzsche shows that there are numerous ways in which we can be wrong about causality. One is when we attribute cause and effect to two things that are always correlated, since, as he claims is often the case, they could both result from a deeper cause. Another error is the one suggested by the above phenomenological account, "that all the antecedents of an act, its causes, were to be sought in consciousness and would be found there once sought…" (TI "The Four Great Errors" 3). In short, this could be called the "*causa sui"* error.[4] A third error, which he entitles "The error of imaginary causes," is similar to the other two. When we have a feeling, Nietzsche says, "We are never satisfied merely to state the fact that we feel this way or that: we admit this fact only—become conscious of it only—when we have furnished some kind of motivation" (TI "The Four Great Errors" 4). So, he says, we

---

ists who argue otherwise. It is important for this paper—which is concerned with determining the type of agency to which Nietzsche is committed—to separate the two premises Leiter mentions and focus on the first one.

[4] Notice that Nietzsche says the error is thinking *all* the antecedents are to be sought in consciousness, not that *any* antecedents are to be sought there. While Leiter may feel this supports what could be considered a secondary thesis—that either way, whether our will is a secondary case, or token-epiphenomenal, we do not have free will—it gives us no argumentative force for his main thesis that Nietzsche views our will as token-epiphenomenal.

invent causes post hoc.[5] The fourth of the Great Errors is what he calls "the error of free will." This fourth error is based on the previous errors and attacks the religious notion of free will associated with being self-caused, which can be seen in his closing remarks on "The Four Great Errors," in which he mentions again that "the mode of being may not be traced back to a *causa prima*" (TI "The Four Great Errors" 8), and therefore we cannot be held responsible in this *religious* sense.[6] Importantly, the fact that Nietzsche observes several ways in which we fail in our causal attributions does *not* mean that he thinks we *always* fail.[7] Leiter points to other textual references, however, in an attempt to support the interpretation that we do always fail. He first cites section 11 of *The Gay Science*, in which Nietzsche argues polemically against the way we have "ridiculously overestimated" our conscious will. Also in that book, Leiter cites evidence from section 333, in which Nietzsche claims that "by far the greatest part of our spirit's activity remains unconscious and unfelt." As Leiter acknowledges, though, these passages do not support the radical thesis that "no conscious belief is part of the causal explanation of *any* action" (Leiter 2009: 119), but he does believe that they, along with the arguments above, debunk the idea that there is a "causal nexus between the conscious experience of will and actions of moral significance" (ibid.). While it may be the case that Nietzsche does not think our conscious will has the causal influence necessary for moral responsibility, his holding this position should not surprise us, since he seems to indicate that the only causal role that could justify moral responsi-

[5] I follow Leiter's interpretation of this third error, which is that this is just an instance of the first error (at least in the way relevant to this discussion) since there is some third, deeper cause that is the cause of both of the events we labeled cause and effect. The specific observation Nietzsche makes with this error is as follows: a feeling arises and since we want to understand *why* it arose, we make up a cause for the feeling (when, in fact, we may not know the actual cause). What this means is that this new causal story was actually itself caused by what, within the causal story, we have understood as the effect.

[6] This is why, for instance, Galen Strawson thinks we do not have "free will." We do not have the type of ultimate buck-stopping responsibility that would justify heaven and hell and since this is what Strawson thinks we mean by "free will," we do not have free will (Strawson 1998).

[7] Further, it seems there is some conceptual room for thinking that we could *always* be wrong with the way we think our conscious causes action, but that it still, in some other way, causes action and is therefore not token-epiphenomenal.

bility would be some sort of *causa sui* ability. So, although this argument further illuminates Nietzsche's view of our conscious will as having a limited (and often inaccurately attributed) causal role, it fails to support Leiter's main thesis that Nietzsche embraces token-epiphenomenalism about conscious willing (as opposed to the view that the will is a secondary cause).

Further, it seems odd that Leiter would cite the passages he does to support the stronger epiphenomenal interpretation. *Gay Science* 11, for instance, seems to point definitively to a causal role for consciousness. In this passage, Nietzsche claims that consciousness "leads to countless errors that lead an animal or man to perish sooner than necessary." Although it is not a positive view of consciousness, this claim certainly commits Nietzsche to a view of its causal efficacy. Further, it is clear in this passage that Nietzsche's problem with consciousness is not that it is epiphenomenal, but the way it has been "overestimated" and taken as "the *kernel* of man." In the end, Nietzsche seems hopeful even about the potential of consciousness. The full version of the sentence quoted by Leiter is: "The ridiculous overestimation and misunderstanding has the very useful consequence that it prevents an all too fast development of consciousness." Nietzsche thinks that consciousness is causally dangerous if it develops too fast, but, as the closing of this section points out, consciousness, correctly developed, is potentitally a quite promising causal agent. The final sentence reads:

> To this day the task of *incorporating* knowledge and making it instinctive is only beginning to dawn on the human eye and is not yet clearly discernible; it is a task that is seen only by those who have comprehended that so far we have incorporated only our errors and that all our consciousness relates to errors. (GS 11)

These remarks on consciousness are admittedly hard to interpret, but in section III, I will give a positive account of consciousness that will bring this and Nietzsche's other references to consciousness into focus.

C. *The Real Genesis of Action*

Lastly, Leiter attempts to draw textual support for his will-epiphenomenalist interpretation of Nietzsche by giving an account of Nietzsche's actual view of the causal genesis of action. The account Leiter offers is related to Nietzsche's broader view of human nature, which Leiter names Nietzsche's "Doctrine of Types" and which he defines as "the doctrine that the psycho-physical facts about a person explain their conscious experience and behavior" (Leiter 2009: 121). The psycho-physical facts Leiter mentions are elsewhere called 'type-facts' and can be understood as the psychological and physiological composition that causes us (unconsciously) to act certain ways. Leiter gives one's "will to power" as an example of a type fact (Leiter 2002: 8), but also speaks more generally about these as physiological facts (such as one's metabolism) and psychological dispositions, also known as drives (Leiter 2002: 71-72). These type-facts cause, but are not equivalent to, one's beliefs and values. Leiter's argument about type-facts relates back to the first "Great Error" discussed above. Because our conscious experience and behavior are always correlated, we assume that one of them (consciousness) caused the other (behavior). What really happens, Leiter suggests, is that these type facts directly cause both our conscious experience and our action (see Fig. 1). While Leiter cites numerous passages in which Nietzsche argues that a non-conscious psycho-physical fact plays an important role in the causal genesis of certain "moral judgments" (Leiter 2009: 116),[8] few of those references warrant the stronger claim that the type fact is the *only* cause, and none of the references are intended to apply to all action. It is compatible with the text Leiter cites that the type-facts are a primary cause (and may even have a great deal of direct influence on our behavior), but that the conscious will is a secondary cause—determined by the type-facts, but able to mediate between the type-fact

[8] BGE 6, 187; GS P2; WP 258; D 119, 542; TI "Skirmishes of an Untimely Man" 37; GM P2, I 15.

and the action in causally relevant ways. So, once again we are left, at the very best, with an argument that simply denies that we have a *causa sui* conscious will.

2. The Interpretive Charity Argument

A. *How the Argument Works*

Just looking at the textual evidence and arguments Leiter gives, we seem to have no reason to think Nietzsche endorses the radical position that our conscious willing never plays a causal role in any of our behavior. Leiter himself seems to acknowledge that there is some ambiguity in Nietzsche's texts on the subject. Faced with this ambiguity, Leiter argues that as a matter of interpretive charity we should read Nietzsche as committed to the view that is "most likely to be correct as a matter of empirical science" (Leiter 2009: 123). To help us see what view this is, he turns to the empirical psychological research of Daniel Wegner and Benjamin Libet, which he presents as uncontroversial evidence that our conscious will is epiphenomenal.[9] In part III of this paper I will argue that Nietzsche's textual references to consciousness are not ambiguous. Nietzsche presents a coherent, causally efficacious account of consciousness and, therefore, we do not need to turn to contemporary empirical research to interpret him charitably. Even with this in mind, however, the research Leiter cites will be important to look at. If it turns out that the research is conclusive, as Leiter claims, then we must conclude that Nietzsche is simply wrong about consciousness—which, of course, should have a significant effect on how we read and understand the rest of his writings. Fortunately, as I will argue in this section, the research Leiter cites is indeed controversial and, therefore, Nietzsche's causally efficacious view of consciousness is not refuted.

[9] That is, epiphenomenal in a way similar to the token-epiphenomenal interpretation for which Leiter argues.

B. *Interpreting the Data*

In discussing the Wegner and Libet experiments, Eddy Nahmias offers a view similar to the token-epiphenomenalism to which Leiter argues Nietzsche is committed. Nahmias calls this view "modular epiphenomenalism" (Nahmias 2010: 348), and it is the view that "non-conscious processes or modules (the 'NC modules') cause the action, while NC modules *also* cause the activity in the C [conscious] modules" (ibid.). If correct, this would mean that conscious willing is an illusion. The question, then, is whether or not the Libet and Wegner experiments actually provide evidence for "modular epiphenomenalism" and for the token-epiphenomenalism reading for which Leiter argues.

Libet's research suggests that the "the brain process to prepare for…voluntary act[s] [begins] about 400 msec. before the appearance of the conscious will to act" (Libet 1999: 51). This conclusion was reached by having a subject perform a "sudden flick of the wrist whenever he/she freely wanted to do so" (50). Whenever this flick occurred, subjects were asked to indicate the position of a rotating dial at the time they were first aware of the wish to move (W). The dial was designed to make a revolution once every 2.56 seconds (making each "second" mark actually represent 43 msec.). During this activity, a technician noted the electrical change in the brain. This electrical change, called the readiness potential (RP) occurred in the brain 350 – 400 msec. before subjects reported they were aware of their wish to move. As Libet notes, however, the conscious awareness, or W value, still occurs 150 msec. before the actual motor act.

It is not surprising that interpretations of Libet's experimental results are debated and controversial. Alfred Mele, responding to the experiment, makes the crucial point that the type of decisions and intentions on which the experiment focuses are the "proximal kind," i.e. 'I will flex my wrist now', and not of the "distal kind," i.e. 'I will work on my paper tomorrow' (Mele 2010:

2). It is possible that when we deliberate and form distal intentions, our consciousness plays a crucial causal role; this type of intention was not tested. A second main criticism is that the RP brain activity could just be an indication of an urge or desire. There is nothing in the Libet experiment that shows an RP necessarily leads to a conscious experience and resulting action. It could be, as Nahmias suggests, that subjects—since they were told to avoid preplanning—formed a conscious distal intention to pay attention to an urge arising within them and to let it go through to action whenever it did arise (2010: 352). The W time then, might simply amount to the subject's awareness of an urge to act and not to his or her awareness of a consciously formed intention. If the RP just represents an urge, then it could be that the conscious awareness of an urge in Libet's subjects 150 milliseconds before the wrist flick simply *allows* the urge to be followed through to action.[10] Nahmias goes even further and argues that "even *if* there are proximal conscious intentions to move and they occur too late to affect the action, it would not follow that *all* relevant conscious mental states were epiphenomenal" (353). As previously mentioned, Nahmias points out that "the subjects' consciously agreeing to move when the urge strikes them surely plays a role in their later actions" (353). To bolster support for this interpretation, Nahmias reflects on his own conscious experience and what he calls "paradigmatic examples of freely willed actions":

> when we drive or play sports or prepare meals, we do not generally form conscious intentions to perform each of the component actions involved in these activities. When we lecture to students or converse with friends, we tend not to think about exactly what we are going to say right before we say it. Rather we may consciously consider what sorts of things we want to say, and then we 'let ourselves go,' though we consciously monitor what's happening and consider how to proceed, for instance, in response to what our students or friends say. (353)

[10] As Nahmias has pointed out, the absence of data concerning cases in which the RPs are not followed by action could be due to the fact that "Libet did not even analyze data on brain activity in cases where subjects felt the urge to flex, but did not flex" (2010: 352).

This experience accords with my own and is equally supported by (or, at the very least, does not conflict with) the Libet experiment.

Given the lively debate and very plausible alternative interpretations of the Libet experiment, it does not look like this contemporary empirical research can support interpreting Nietzsche as a token-epiphenomenalist for reasons of philosophical charity, contrary to what Leiter argues. Still, Leiter also cites the work of Daniel Wegner to support this interpretation, so we must also consider his research.

I see no reason to doubt that Wegner's research supports and gives empirical verification to the arguments Nietzsche makes regarding falsely attributed causes. Both Wegner's research and Nietzsche's texts show how our phenomenology of willing can be separated in many instances from actual causation. Wegner, however, seems to make a further inference from his data—namely, that our phenomenology of willing is *always* cut off from the actual cause of our actions. This alternative picture Wegner gives is quoted by Leiter as follows: "Unconscious and inscrutable mechanisms create both conscious thought about action and the action, and also produce the sense of will we experience by perceiving the thought as cause of the action" (Leiter 2009: 122). This interpretation is strikingly similar to the token-epiphenomenalist view for which Leiter argues. Before we accept Wegner's interpretation, though, we must take a closer look at Wegner's research.

Wegner states that there are three main conditions for our experience of conscious will: the conscious thought has to be consistent with an action, prior to it, and exclusive to it (2008). Wegner cites multiple studies that test how the meeting of each of these conditions—consistency, priority, and exclusivity—can lead to your feeling of being a causal agent when in

fact you are not one. I will review one of the studies focused on consistency that I feel is representative of the other studies.

In what I will call the Voodoo Experiment, subjects are told they are participating in an experiment to test psychosomatic influences on health. The subject plays the role of a witch doctor sticking pins in a doll, while someone (who is secretly in on the experiment) plays the victim. Some subjects had victims that arrived late and were impolite, while other subjects had "normal and likeable" (2008: 229) victims. The subjects who were led to dislike the victim beforehand were more likely to believe their voodoo curse caused the victims headache than those subjects who had normal and likeable victims. The same increase occurred when subjects were instructed simply to "'think negative thoughts' about the victim" (ibid.). Importantly, as Wegner notes, participants were given an article suggesting that you could psychosomatically influence another's health (ibid.). It is hard to determine how these results should be interpreted. It does show that consistency is an important part of how we determine our own causal influence, but it does not show that such determinations of causal inference are wrong. Subjects were *wrong* in the experiment, but that is only because they were deliberately tricked. Consider the following situation: you send an email to your friend encouraging her to go to a talk on campus the next day. You are pretty she will not know about the talk on her own, as it was not advertised. The next day you see her at the talk and assume she read your email. Yet it turns out that she never checked her email, but just happened to be walking by and stopped in to hear the talk. You considered yourself the cause of her appearing in the lecture hall, when in fact, since she never checked her email, there was no way you could have been. This is analogous to the Voodoo experiment, and we should be no less worried about the Voodoo Experiment's conclusion than we

are about this more familiar type of mistake.[11] The experiment, therefore, is far from conclusive evidence about our failure to attribute causes accurately in all or even normal circumstances.

Beyond his data, Wegner appeals to other arguments (including a citation of Libet, which I have already discussed above) that he feels justify the leap to the stronger interpretation, which is the interpretation Leiter uses to support his token-epiphenomenalist reading of Nietzsche. For my purposes, though, it is his data that is of interest. The data is the "empirical research" that, along with Libet's, is supposed to shift interpretative charity in favor of Leiter's thesis that Nietzsche is a token-epiphenomenalist. All this data seems to show, however, is that just as Nietzsche's "Four Great Errors" suggest, we frequently inflate or simply wrongly attribute our own causal influence. Showing we sometimes make these mistakes does not, of course, show that we always do. Nahmias draws a similar conclusion and suggests that the best interpretation of data such as Wegner's is one that draws an analogy with visual illusions (2010: 351). Nahmias thinks that just as visual illusions that deceive us do not demonstrate that all of our visual experiences are mistaken, we should not take these causal illusions to demonstrate that all of our causal attributions are mistaken. Nahmias takes the analogy a step further. He writes, "Indeed, as with visual illusions, explanations for illusions of will may be offered in terms of a generally *reliable* system, which sometimes produces inaccurate output because of some unusual feature of the situation" (ibid.), and he points out that all of the cases Wegner cites are unusual in some way.

[11] The apparent force of the Voodoo Experiment, and a difference between it and my example, is just how unbelievable we think the causal attribution is in the Voodoo Experiment. But, again, subjects were prepared with an article suggesting that this causal mechanism was real and then given a clear correlation (which they must have thought was unlikely to be just a coincidence). Also, the subjects did not have any reason to consider that such causal influence was unbelievable, since most people, I assume, had not read articles opposing such a possibility and did not have well thought-out and thoroughly researched stances on psychosomatic influence.

After reviewing the empirical research to which Leiter appeals, it is apparent that the data offers no conclusive evidence for a token-epiphenomenalist reading. Both the Libet and Wegner data are compatible, at least on some plausible interpretations, with consciousness playing *some* causally efficacious role in action. Since, as I have shown, the textual arguments that Leiter gives are also insufficient to motivate a token-epiphenomenalist reading, it seems that if we want to give a charitable interpretation of Nietzsche's texts, it would be best to read Nietzsche as attributing at least *some* causal role to consciousness rather than the "grossly implausible" (Graham 1993: 185) view of epiphenomenalism. Still, it is not hard to imagine Nietzsche holding a *radical* view that many intelligent people might find "grossly implausible," and so leaving this important interpretative question up to what is most charitable or simply withholding judgment on the issue may seem rather unsatisfying. In the rest of the paper, I will look more closely at Nietzsche's textual references to consciousness and show that we do not need to rest on an argument from interpretive charity to determine how we are going to understand Nietzsche's view of consciousness. There is one coherent view of consciousness that Nietzsche unambiguously presents. Far from being epiphenomenal, this view maintains that while consciousness has been "ridiculously overestimated," it still can causally influence action.

III. CONSCIOUSNESS AS A NON-ESSENTIAL PROPERTY

The nature of our consciousness, for Nietzsche as for most philosophers, is something that distinguishes human animals from others. It is not, however, a distinct "Cartesian" faculty. The Cartesian division of mind, to which Nietzsche thinks many still adhere, even if only implicitly, has no grounding. For instance, in one of the many passages driving home this point, Nietzsche draws an analogy with lightning:

> For just as the popular mind separates the lightning from its flash and takes the latter for an *action*, for the operation of a subject called lightning, so popular morality also separates strength from expressions of strength, as if there were a neutral substratum behind the strong man, which was *free* to express strength or not to do so. But there is no such substratum; there is no "being" behind doing, effecting, becoming; "the doer" is merely a fiction added to the deed—the deed is everything. (GM I 13)

The same sentiment is at work in what is probably Nietzsche's most oft-recognized passage addressing "freedom of the will," *Beyond Good and Evil* 21. What Nietzsche is attacking is not every notion attached to freedom or attached to the will, but the idea that there is some separate, completely free, *doer*, which is self-caused and chooses in an undetermined way to perform actions. In his words, the mistake is in desiring "'freedom of the will' in the superlative metaphysical sense" (BGE 21).

Understanding this criticism of the idea of a self-caused consciousness is critical to understanding Nietzsche's positive view of consciousness. What has complicated the debate about Nietzsche's views on consciousness is that he seems to offer contradictory statements regarding the causal efficacy of consciousness. At times he seems to say it has no role at all, and other times he straightforwardly commits himself to the causal efficacy of consciousness. These disparate references can be accounted for, however, as long as we are careful about the distinction between Nietzsche's references to a faculty of consciousness sitting behind all thought and action and a non-dualistic conception of consciousness. Nietzsche certainly thinks the idea of consciousness as a non-physical *causa sui* faculty is wrong, and if someone has that picture in mind and thinks that such a faculty is what is causing action, he is indeed dealing with an epiphenomenon (since that faculty does not exist, of course, it is not causally efficacious). Arguing for this claim is often what Nietzsche is doing in the passages cited by Leiter (as I have shown in section II.1) and in other passages in which he denies the causal efficacy of consciousness. The

positive references to consciousness, on the other hand, refer to his own view of consciousness which does not posit it as a faculty but rather as a *non-essential property* of mental states.

1. *Genealogy of Morals* II 16

To begin to see the positive account of consciousness given by Nietzsche, we can turn to two important passages: *The Genealogy of Morals* (hereafter 'GM') II 16 and *The Gay Science* (hereafter 'GS') 354. I will begin by outlining the hypothesis given in GM II 16, and then, dealing with GS 354 and interpretive work done by Paul Katsafanas, I will give a theory of mind that will allow GM II 16 to come into focus and clarify the important role consciousness plays in Nietzsche's work.

The focus (and title) of the second essay of GM is "'Guilt,' 'Bad Conscience,' and the Like." Section 16 is the pivotal section of the second essay, in which Nietzsche introduces his "hypothesis concerning the origin of the 'bad conscience'." What Nietzsche means exactly by "the bad conscience" is something that will become clearer as we work through this passage. The short version of the bad conscience hypothesis is as follows:

> I regard the bad conscience as the serious illness that man was bound to contract under the stress of the most fundamental change he ever experienced—that change which occurred when he found himself finally enclosed within the walls of society and of peace.[12]

Giving an unmistakably evolutionary account, Nietzsche compares the situation of pre-"bad conscience" human animals to sea animals forced to become land animals in order to survive. According to Nietzsche, the human animal was well adapted to "wilderness, to war, to prowling, to adventure" when suddenly, forced to live together in civilizations, "their instincts were disvalued and 'suspended'." Here, tellingly, Nietzsche makes a clear distinction between consciousness and the unconscious:

[12] All lengthy quotes in this section will be from GM II 16 unless otherwise noted.

> [These early human animals] felt unable to cope with the simplest undertakings; in this new world they no longer possessed their former guides, their regulating, unconscious and infallible drives: they were reduced to thinking, inferring, reckoning, co-ordinating cause and effect, these unfortunate creatures; they were reduced to their 'consciousness,' their weakest and most fallible organ!

The aggressive instincts for war, prowling and adventure, described above, no longer had a value or place in this new group-living situation, but that does not mean they immediately ceased to exist. Our instincts, on Nietzsche's view, are not the type of things that can simply be extirpated. Instead, these instincts can only be redirected. In society, when man became a social animal, the outward expression of the aggressive instincts were forcibly blocked. Then, in a pro-to-Freudian insight, Nietzsche claims that, with nowhere to go, these instincts turned inward:

> All instincts that do not discharge themselves outwardly *turn inward*—this is what I call the *internalization* of man: thus it was that man first developed what was later called his 'soul'. The entire inner world, originally as thin as if it were stretched between two membranes, expanded and extended itself, acquired depth, breadth, and height, in the same measure as outward discharge was *inhibited*.

Nietzsche's greatest insight into the nature of our mental life—to which I will return throughout my paper—is that the development of our entire inner world, including consciousness, was necessarily connected with our transition into social living situations. How this transition took place is touched briefly upon in section 16, but is discussed in more detail in GM II 3. In that section, Nietzsche describes a brutal process of memories being burnt into the mind. It is not hard to imagine the type of story Nietzsche has in mind here. Man, either by force, or simply by necessity (Nietzsche mostly believes this was by force), enters into a political organization in which certain acts, such as attacking another human, are not tolerated. The transgressor is then brutally punished every time he does this act, and through "blood, torture, and sacrifice" (GM II 3) a memory is created so that he can *stop* his aggressive instincts from expressing themselves outwardly. So, Nietzsche says, the origin of the "bad conscience" is: "Hostility, cruelty, joy in persecuting, in attacking, in change, in destruction—all this turned against the possessors of such

instincts" (GM II 16). The importance of this hypothesis for Nietzsche's view of human history cannot be overstated. He claims that with this development the "instincts of wild, free, prowling man turned backward *against man himself*" and began "the gravest and uncanniest illness, from which humanity has not yet recovered, man's suffering *of man, of himself*—the result of a forcible sundering from his animal past, as it were a leap and plunge into new surroundings and conditions of existence" (ibid.).

But, as is common in Nietzsche, he pairs this dismal view of the development of "bad conscience," with a positive outlook, full of possibilities:

> Let us add at once that, on the other hand, the existence on earth of an animal soul turned against itself, taking sides against itself, was something so new, profound, unheard of, enigmatic, contradictory, *and pregnant with a future* that the aspect of the earth was essentially altered.

Nietzsche ends the section with unbridled enthusiasm:

> From now on, man is *included* among the most unexpected and exciting lucky throws in the dice game of Heraclitus' "great child," be he called Zeus or chance; he gives rise to an interest, a tension, a hope, almost a certainty, as if with him something were announcing and preparing itself, as if man were not a goal but only a way, an episode, a bridge, a great promise.—

The language of these closing remarks is unmistakably similar to the enthusiastic proclamations made by Nietzsche in important passages of his other major works, such as the preface to *Beyond Good and Evil* and throughout *Thus Spoke Zarathustra.*[13] While the details are still unclear, GM II 16 tells the story of pre-civilized man becoming a social animal and explains how this development fundamentally changed what it meant to be a human. This fundamental change was the creation of an entire inner world, which at first is a grave illness, but which also makes man interesting and "pregnant with a future."[14]

---

[13] See for instance "The Gift Giving Virtue" from part one of *Thus Spoke Zarathustra*.

[14] At the start of GM II 19, he uses this same imagery to reiterate the view of the bad conscience and inner world as both an illness and a source of hope: "The bad conscience is an illness, there is no doubt about that, but an illness as pregnancy is an illness."

GM II 16 and my presentation of it raise a number of issues and questions that demand further clarification. For now, however, I will stress only two important points made in this passage.

First, what has been largely underappreciated in Nietzsche scholarship is the explicit distinction made here by Nietzsche between the unconscious and consciousness. Early man—man before the invention of bad conscience—survived and flourished by means of *unconscious* mechanisms. The change that took place was "the most fundamental change" because it required humans to develop and rely on their consciousness after the unconscious mechanisms were no longer effective. Nietzsche calls humans' consciousness "their weakest and most fallible organ," and it is presumably part of what he describes as the expanding "inner world" that was originally "as thin as if it were stretched between two membranes."

This section on its own should be enough to cast suspicion on the view that consciousness is epiphenomenal: we could not be forced to rely on something that can have no causal influence on our actions. Still, this section is only a rough overview of Nietzsche's view, and will require further explication to understand exactly what differentiates consciousness from the unconscious, how consciousness is "an organ," and how it, with the rest of this inner world, expanded. A few initial things should be cleared up before I discuss the second important initial point about GM II 16. First, as I will discuss again in section III.3, consciousness is not equivalent to the "bad conscience" and does not make up the inner world in its entirety, but it is intricately related to "bad conscience" and its proliferation. Secondly, although he refers to consciousness as an "organ," this should not be taken to mean Nietzsche thinks consciousness is a wholly separate faculty. Thirdly, what the "organ" and "membrane" language do show is that the account Nietzsche gives us is one that can be understood *biologically*. Consciousness, and

this whole expanding inner world, could, potentially, be located physically.[15] In section III.2, I will explore in more detail Nietzsche's positive, non-dualistic conception of consciousness.

The second important initial point I will make about GM II 16 is that Nietzsche refers to the change that occurs in man becoming a social animal as "the most fundamental change" the human animal has ever experienced. What is lost in Leiter's epiphenomenalist interpretation is that in important ways humans are different from animals. Yes, we are indeed animals, and Nietzsche is keen on reminding his readers of this point, but we are not just *any other* animal. Maudemarie Clark and David Dudrick make this same point against Leiter's naturalist reading, saying that, in an important way, Nietzsche views "human thought and action" differently from the thought and action of other animals (Clark and Dudrick 2009: 248).[16] Leiter's interpretation, which claims that unconscious type-facts do *all* the explaining, is dangerously close to sounding as though we are, by and large, no different from other animals and that the one thing that does seem to distinguish us, our consciousness, is not important or causally efficacious at all. It is important to note that the subjects Nietzsche is concerned with are *human* animals, the "valuing animal as such" (GM II 8). Their values, ideals, and "entire inner world" are what Nietzsche considers unique, "unheard of," "exciting," "new." While it may be possible in theory to have all of these distinctively human features be a part of a completely unconscious "bad conscience," that reading is unmotivated, would not capture the strength of this uniqueness, and, as I will show in this paper, it is not Nietzsche's view.

While GM II 16 gives us a hypothesis that outlines in general how this "entire inner world" was expanded, it leaves important questions about the nature of this inner world unans-

[15] Or, put differently, consciousness is not a different *kind* of thing compared to the rest of our body.
[16] Clark and Dudrick, then, argue that in some important sense Nietzsche is not a strict naturalist. This claim need not follow, however, and the rest of my paper will show how this difference between human animals and other animals can be preserved, while still maintaining that Nietzsche is a naturalist about human thought and action.

wered. If consciousness is not a separate faculty, how *exactly* did it "come to be" within this one particular animal and what is its connection to the rest of the "inner world"? With the above framework in mind we can turn to the answer Nietzsche gives in GS 354, which offers a parallel hypothesis about the origin of consciousness specifically.

2. *Gay Science* 354

The section entitled "On the 'genius of the species'" asserts at the beginning that:

> we could think, feel, will, and remember, and we could also 'act' in every sense of that word, and yet none of all this would have to 'enter our consciousness' (as one says metaphorically).... Even now, for that matter, by far the greatest portion of our life actually takes place without this mirror effect; and this is true even of our thinking, feeling, and willing life....

At first this seems like an outlandish claim. It may be conceivable how we could *feel* unconsciously, in some sense *will* unconsciously, and maybe in some sense *remember* unconsciously, but *thinking* unconsciously seems like a blatant contradiction in terms. To attempt to understand this initially startling claim, I will refer to recent work done by Paul Katsafanas on Nietzsche's theory of mind. One thing that Katsafanas notes, and that is apparent in the above quotation, is that for Nietzsche consciousness is not an *essential* property of mental states. While it is implied in the above quote that mental states such as thinking, feeling, and willing can be conscious, Nietzsche says these same mental states can occur without the property of consciousness.[17] Following the passage I just quoted above, Nietzsche poses a question that naturally arises from this claim: "*For what purpose,* then, any consciousness at all when it is in the main *superfluous*?" (GS 354). In the rest of the section, Nietzsche gives us his answer and "the perhaps extravagant surmise that it involves" (ibid.). The observation that drives his insight into consciousness is that

[17] This is also why Nietzsche puts "enter our consciousness" in quotes and remarks that this is "what one says metaphorically." To say that something could "enter consciousness" is to assume that consciousness is a substantive faculty, or an empty room waiting to receive content. This is certainly not Nietzsche's view. As a non-essential property of certain mental states, it does not make sense literally to assume something can "enter our consciousness." Some mental states just are conscious, and others are not.

"the subtlety and strength of consciousness always were proportionate to a man's (or animal's) *capacity for communication,* and as if this capacity in turn were proportionate to the *need for communication*" (ibid.). So his answer is ultimately that "*consciousness has developed only under the pressure of the need for communication*" (ibid.).

What follows from this claim echoes the point made above about the nonessential nature of consciousness: "Consciousness is really only a net of communication between human beings; it is only as such that it had to develop; a solitary human being who lived like a beast of prey would not have needed it" (GS 354). Humans can, and did, get by without consciousness. The fact that Nietzsche here mentions pre-civilized man's not needing consciousness I take to be significant and relate it to the story told in GM II 16, in which Nietzsche states that consciousness was part of the inner world that the process of civilization caused to expand. Before the development of civilization and the creation of the bad conscience, man, the beast of prey, did not need consciousness. The connection to the story in *The Genealogy of Morals* is made more explicit in the sentences that immediately follow in GS 354:

> That our actions, thoughts, feelings, and movements enter our own consciousness—at least a part of them—that is the result of a "must" that for a terribly long time lorded it over man. As the most endangered animal, he *needed* help and protection, he needed his peers, he had to learn to express his distress and to make himself understood; and for all of this he needed "consciousness" first of all, he needed to "know" himself what distressed him, he needed to "know" how he felt, he needed to "know" what he thought.

It was when humans needed to live with each other, as already discussed in GM II 16, that consciousness developed. However, this passage is more than just a reiteration of the points made in GM II 16. Nietzsche explicitly connects man's need to "express distress and to make himself understood" to consciousness. This explains why consciousness "is only a net of communication" and only develops as such.

So what, then, distinguishes mental states with the property of consciousness from mental states that do not have that property? Nietzsche seems to suggest that it is the use of words: "The thinking that rises to *consciousness* is only the smallest part of all this—the most superficial and worst part—for only this conscious thinking *takes the form of words, which is to say signs of communication*, and this fact uncovers the origin of consciousness." So the origin and nature of consciousness is betrayed in the fact that conscious thinking occurs only in words. A mental state with consciousness, is therefore, one that is expressed in words. For Nietzsche, this means that they can be conceptually articulated—"words are acoustical signs for concepts" (BGE 268).[18]

At this point it will help to put all of this in perspective by considering some distinctions made by Katsafanas. As I have already pointed out, for Nietzsche, all of our mental states are either conscious or unconscious, and the difference between these two types of states is that conscious states are conceptually articulated, while unconscious states are non-conceptually articulated (they are articulated, in other words, just not "conceptually articulated"). One noteworthy aspect of this distinction is that the difference between conscious and unconscious mental states is *not* a difference of awareness, which is often considered the distinguishing feature of consciousness. I will discuss objections to and concerns about this aspect of Nietzsche's view below. For now, it will be helpful simply to notice that, for Nietzsche, there are unconscious perceptions, and thus we may be in some sense aware of our unconscious mental states as well.

One thing that Katsafanas concludes by showing how Nietzsche was influenced by aspects of both F.A. Lange's and Schopenhauer's views on mental states, is that unconscious perceptions are articulated and have form, though this form is not conceptually articulated (2005: 5). One defining difference between these two types of articulated perception—conscious and un-

[18] For an early mention of this idea see *Daybreak* 257.

conscious—is the classifying awareness that is a part of conscious perception. Thus, Katsafanas claims that for Nietzsche:

> [conscious] perceptions involve a classifying awareness, which presents objects as instances of concepts that the perceiver can employ in abstract thought...whereas unconscious perceptions involve only a discriminatory ability, only a perceptual sensitivity to features of the environment....(2005: 9)

Thus, unconscious perceptions cannot be employed in abstract thought. Katsafanas tries to clarify just what this distinction means, and I will discuss his example below, but it should be noted that unconscious perception is, by the definition given, impossible to express in words.

The transition from an unconscious mental state to a conscious mental state is "conceptualization," which Katsafanas explains as follows: "Perceptual content would be conceptualized if the perceived object were represented as an instance of some concept, that is, as a token of some type" (2005: 7). As Katsafanas notes, Nietzsche himself gives an example of this process in section 192 of *Beyond Good and Evil*:

> Our eye finds it more comfortable to respond to a given stimulus by reproducing once more an image that it has produced many times before, instead of registering what is new and different....[We] do not see a tree exactly and completely with reference to leaves, twigs, color, and form; it is so very much easier for us simply to improvise an approximation of a tree.

So, when we conceptualize, we have an unconscious perception before us and, automatically, we translate some aspect of this into a fixed concept. If we have a concept already that resembles that articulated unconscious perception, we will translate it as that concept, rather than producing a new concept. Once that unconscious perception has been translated into a concept, TREE, we have a word that we can now employ to express the perception (or, more properly, an aspect of that perception) to other humans.

A. *Initial objections and response*

While the preceding remarks regarding language and "conceptualization" seem plausible, one might object that this simply cannot be *all* consciousness is. Someone might, for instance,

claim that he can be aware of something without being able to articulate it conceptually. Katsafanas answers this objection by pointing out that Nietzsche uses the word "consciousness" in a different way than it is commonly understood (at least in contemporary English) (2005: 14). As I have already stated, Nietzsche does not use awareness to draw the line between consciousness and unconsciousness. This point follows from his rejection of consciousness as a substantive faculty. If there is no substantive faculty, no *doer*, behind deeds, there is no subject that can have this awareness. A distinction grounded on the claim of awareness in this dualist sense would have no meaning.[19] Thus, for Nietzsche a mental state of which you are "aware," but which you cannot articulate conceptually is just what he calls an unconscious mental state.

While the first objection may be answered, the answer seems to bring up a different objection; namely, why call this distinction one between *conscious* states and *unconscious* states at all? Katsafanas offers two reasons why Nietzsche does this. First, the ability to articulate perception conceptually *is* what distinguishes humans from animals, and so "if we are looking for a distinction in the mental, then, the natural place to draw it is between conceptually and nonconceptually articulated states" (2005: 14). Secondly, "conscious states," under this definition, "are *accessible* to us in a way that unconscious states are not; in particular, they are communicable" (ibid.).

If the reasons presented by Katsafanas still seem unsatisfactory, it will help to consider just how ingrained our Cartesian view of the mind is. Nietzsche never tires of emphasizing how embedded in our language (and thus, in our thought) is this dualist view. For example in *Beyond Good and Evil* 16 and 17 (also cited by Leiter), Nietzsche attacks the idea that "I think" is an

[19] This is not to say there cannot be drives or instincts that we are not aware of or are not always aware of. The point is that using "aware" at all, confuses the problem as it seems to refer to an observing, separate faculty that *sees* certain states and not others. So, I take it that many of our unconscious perceptions are, what we would commonly consider to be mental states we are "unaware" of, but that is only because of the certain ways our conceptually articulated perception has been closed off from them.

immediate certainty, saying that even "it thinks" goes too far and "does not belong to the process itself," and saying that "more rigorous minds" will perhaps "get along without the little 'it' (which is all that is left of the honest little old ego)." The confusion, Nietzsche claims, is due to "grammatical habit," and he forcefully ends the first paragraph of 16 by saying: "I shall repeat a hundred times; we really ought to free ourselves from the seduction of words." The idea of a Cartesian ego that can sit above and be aware or not be aware of certain mental states, is *not* what you are going to get—nor is it what you should expect—in Nietzsche's account of consciousness.

The above point is analogous to a point made in the free will debate. A free will skeptic like Galen Strawson, for instance, thinks that what "free will" refers to is a *causa sui* ability, in which one has buck-stopping responsibility, and since there is no way that we can have this ability, there is no way that we can have free will. A common compatibilist response is to acknowledge that *this* conception of free will is incoherent, or even impossible, and then to give a definition of free will that does fit with determinism, either by making a different distinction or by clarifying what the distinction really was that we were making. Similarly, Nietzsche offers a different hypothesis about what distinguishes conscious mental states from other mental states, since he considers it untenable that the distinction should hinge on the distinction between 'awareness' and 'unawareness'.

Before looking at the process of conceptualization in closer detail, it will help to pause and consider, once again, the comments Nietzsche makes about being able to "reason" or "think" unconsciously. If we understand this passage as referring to the standard meaning of the word "unconscious," in which the word is basically equivalent to "unaware," then we simply cannot make any sense of what Nietzsche is talking about. So either he is not using the term with this

meaning in mind, or he is simply not making any sense. As I have argued, there is good evidence to suggest he has a different meaning in mind, namely that "unconscious" refers to mental states that are non-conceptually articulated. Thus, we *are* still aware of them, and they *are* still articulated. So, another advantage of the view I have presented so far is that we have at least a starting point for understanding how Nietzsche can claim we can "think" and "reason" unconsciously, so we need not disregard these comments as nonsensical.

B. Conceptualization

With Katsafanas' definitions and clarifications in mind, we can return to GS 354 and see how they are grounded in that passage. Nietzsche claims, "The emergence of our sense impressions into our own consciousness, the ability to fix them and, as it were, exhibit them externally, increased proportionately with the need to communicate them to *others* by means of signs." Here Nietzsche is relating our ability to conceptualize, that is, to fix our sense impressions in such a way that they can be exhibited externally, directly to the need to communicate them to others. Before we needed to communicate with others, we never had a need to conceptualize our "sense impressions." When we came across a tree, we could think about it, have feelings about it, remember it, and even, Nietzsche claims, reason about it, but none of this would be done using concepts, the "signs of communication." Then there arose a need to communicate, and we had to develop our ability to fix our sense impressions so that they could be understood by others.

It will be helpful here to consider a few initial worries about the accuracy of these connections, from sense impression to conceptualization on the one hand, and from conceptualization to communication on the other. First, when I have a conscious mental perception TREE, I have lost a lot of the intricate detail that was contained in my unique unconscious mental perception of that particular tree and its surroundings. Secondly, when we express this conceptualized

perception to another person, the perception that is evoked in the other's mind (i.e., the conceptual perception TREE which he has conceptualized from his own unique unconscious perceptions of trees) is not going to be the same conceptual perception I have, but rather *his* concept TREE. Of course, if one is able to conceptualize multiple parts of this tree and *bring them along*, as it were, one could have a more detailed and accurate conceptual perception of TREE. Also, as one uses his concepts, such as TREE, in conversation with us, we can imagine an increased accuracy in his ability to communicate that content, since we would be refining our concept to match the intricacies in his concept. There is good evidence, however, that Nietzsche thinks this hopeful possibility (i.e., that using our concepts will yield greater accuracy) is *not* actually what happens when we communicate with others.[20] Even if these initial worries did dissipate over time, though, there is something else that limits the accuracy of our concepts (and, therefore, our consciousness), and this is Nietzsche's most powerful insight in GS 354.

Nietzsche voices the main idea of GS 354 about halfway through the passage. "My idea," he says:

> is...that consciousness does not really belong to man's individual existence but rather to his social or herd nature; that, as follows from this, it has developed subtlety only insofar as this is required by social or herd utility. Consequently, given the best will in the world to understand ourselves as individually as possible, "to know ourselves," each of us will always succeed in becoming conscious only of what is not individual but "average."

[20] While these points are all made in Nietzsche's later works (*The Gay Science* and *The Genealogy of Morals*), from which I have been quoting, similar and even more radical points are made in a much earlier unpublished fragment titled, "On Truth and Lie in an Extra Moral Sense." There, Nietzsche claims that:

> Every word immediately becomes a concept, inasmuch as it is not intended to serve as a reminder of the unique and wholly individualized original experience to which it owes its birth, but must at the same time fit innumerable, more or less similar cases—which means, strictly speaking, never equal—in other words, a lot of unequal cases. Every concept originates through our equating what is unequal. (TL, p. 46)

Nietzsche goes on to claim that eventually we forget that there are distinctions among our individual experiences of some "concept," and this forgetting gives rise to the idea that "in nature there might be something besides the leaves which would be 'leaf'" (ibid.). This leads Nietzsche to the conclusion that "truths are illusions about which one has forgotten that this is what they are" (ibid.).

So the overarching problem with consciousness is that the force driving conceptualization, at least initially, is "social or herd utility." Why and how an unconscious perception was originally conceptualized has to do with the use that perception had for some social end. Further, the way it is received and used by other humans is also determined by the necessity of the social situation. We can distinguish two separate points here: one about *how* perceptions are conceptualized, and one about *what* part of a perception is conceptualized.

C. *How Unconscious Perceptions are Conceptualized*

First, accuracy in what we do conceptualize is not the goal in our conceptualization or in our communication of the concept. Nietzsche, of course, accounts for the fact that in certain contexts, such as philosophy, accuracy is what drives some conceptualizations.[21] Katsafanas makes a similar point by citing GS 111. Here Nietzsche gives us another evolutionary account, but this time, he presents a concept fundamental to logical thought, namely 'substance':

> In order that the concept of substance could originate—which is indispensible for logic although in the strictest sense nothing real corresponds to it—it was likewise necessary that for a long time one did not see nor perceive the changes in things. The beings that did not see so precisely had an advantage over those that saw everything "in flux." At bottom, every high degree of caution in making inferences and every skeptical tendency constitute a great danger for life. No living beings would have survived if the opposite tendency—to affirm rather than suspend judgment, to err and *make up* things rather than wait, to assent rather than negate, to pass judgment rather than be just—had not been bred to the point where it became extraordinarily strong. (GS 111)

The idea is that it was evolutionarily beneficial to develop the capacity to conceptualize *quickly* and decisively, without a great deal of concern for accuracy. If something is conceptualized in such a way as to serve its social purpose, then that is good enough; there was nothing pushing the conceptualization towards accuracy. In fact, as Nietzsche seems to point out with respect to the

[21] Explaining how to make sense of conceptualizations that are driven by accuracy is not necessary for the purposes of this paper. With that said, in section II.3, I lay the groundwork for understanding *how* this could be possible. In short, while our conceptualizations are initially determined by our social drives, other drives can take precedence in determining conceptualizations. An example of such a drive is Nietzsche's "will to knowledge." However, this "will to knowledge" is not always a good thing. It represents a "*faith*" in unconditioned truth (GM III 24), and Nietzsche thinks that such non-perspectival ideas have no warrant, are part of what he labels the "ascetic ideals," and are responsible for turning man into "*the* sick animal" (GM III 13).

concept ‘substance’, the conceptualization could serve its social purpose *because* it did not strive for accuracy and *because* it in fact misrepresented the unconscious perception (by seeing things as fixed when in reality everything is in flux). So this, then, is the first point stated somewhat differently: survival of the species is what drives conceptualization (at least initially), and conceptualizations that are quick, decisive, general, and efficient enough will be more beneficial for survival than conceptualizations that strive towards accuracy. Nietzsche closes the section just quoted, about the conceptualization of substance, by saying: “We generally experience only the result of this struggle because this primeval mechanism now runs its course so quickly and is so well concealed” (GS 111). In other words, we have an automatic conceptualizing mechanism within us that pushes our conceptualizations not toward accuracy, but toward speed and efficiency.

At this point, it might be objected that, over time, these two different, often competing goals of efficiency and accuracy would actually converge. So, if we accept that reality is objective (or mind-independent), then, if our concepts fail to represent it accurately, then these concepts will eventually fail to serve their purpose. For example, if I conceptualize an unconscious perception of fire (and, therefore, turn it into a conscious perception) without including some notion of ‘extreme heat’, I will eventually—in identifying other perceptions of fire—be forced to understand FIRE as including extreme heat. Thus, potentially, the continued experience with and use of our concepts over time should eventually eliminate most of our inaccurate concepts. While I think Nietzsche must agree that this could happen to at least a small degree within a certain worldview, he, as I will argue below, is also committed to a view that there are numerous philosophical worldviews that are mutually exclusive, but internally coherent. Katsafanas describes Nietzsche’s view about this by adopting a metaphor from Quine: “conceptual frameworks

are responsive to experience only at their edges; there are many different, mutually incompatible ways of cashing out the content of unconscious experience" (Katsafanas 2005: 17). Textual evidence in Nietzsche for this Quinian point can be found in section 20 of *Beyond Good and Evil*. Here, Nietzsche attempts to explain how worldviews that are similar to each other are the result of a few foundational unconscious perceptions being conceptualized similarly, and worldviews that are different from one another can be traced back to a few foundational unconscious perceptions being conceptualized differently. A key part of the passage reads:

> Where there is affinity of languages, it cannot fail, owing to the common philosophy of grammar...that everything is prepared at the outset for a similar development and sequence of philosophical systems; just as the way seems barred against certain other possibilities of world interpretation. It is highly probable that philosophers within the domain of the Ural-Altaic languages (where the concept of the subject is least developed) look otherwise "into the world," and will be found on paths of thought different from those of the Indo-Germanic peoples and the Muslims: the spell of certain grammatical functions is ultimately also the spell of *physiological* valuations and racial conditions. (BGE 20)

So, simply based on how certain grammatical concepts were conceptualized, and thus consciously understood, entire belief systems and valuations formed and *had* to form in a certain way. One of the most important conceptualizations that takes place in this regard is the one that Nietzsche uses as an example: one's conceptualization of the subject "I" and the general notion of a "subject." It is not insignificant that the section that follows *Beyond Good and Evil* 20 is Nietzsche's vociferous attack on "*causa sui*" free will. It is because of our strong, well-developed concept of the subject, which desires "to bear the entire and ultimate responsibility for one's actions oneself" and "to pull oneself up into existence by the hair, out of the swamps of nothingness" that we have developed this "monstrous conception" of free will (BGE 21). It is the same reason we have developed a conceptualization of "consciousness" as a separate faculty and that we attribute to it powers that it does not have.

Another consideration involves the type of content being conceptualized. While there are certain facts about the world that seem non-negotiable and that will therefore come to have a refined conceptualization that is similar for all languages, some content can be conceptualized in a range of different ways. For instance, that a certain burning sensation is connected with the conceptualization FIRE and that the concept is deeply connected with "extreme heat" are going to be broadly consistent, while other concepts, such as those involved in value judgments, have a nature that allows for a considerably greater amount of *leniency* in their conceptualization. These latter conceptualizations, especially those dealing with value, are Nietzsche's primary concern. These value judgments are by and large what are involved when one philosophizes, and under this interpretation are, as I will show, central to Nietzsche's main task of the "revaluation of all values." Before I discuss this, however, it will be important to understand a second worry about the way conceptualization occurs.

D. *What content is conceptualized*

In the preceding section I have shown how the process of conceptualization is not motivated by a desire for accuracy. Rather, the process of conceptualization is motivated by imminent social need. The upshot of this point is that conceptualizations that are quick, rigidly fixed, and good enough will win (and have won) out over conceptualizations that are slow, sensitive to the complexities and fluidity of a perception, and that strive for accuracy. Further, I have shown that—especially with the concepts that matter most to Nietzsche—this is not a worry that will necessarily dissipate with experience. So this first worry is about *how* an unconscious perception is conceptualized. Another separate, but important, worry about this conceptualization is *what* content is chosen for conceptualization. Since consciousness, that is our conceptually articulated mental states, have "*developed only under the pressure of the need for communication*" and,

therefore, "developed subtlety only insofar as this is required by social or herd utility" (GS 354), what has been picked out for conceptualization has always been only that part of unconscious perception that was *needed*. At least initially, there was no motivation to conceptualize something unless it was required to communicate to others.

First, a simplified, concrete example of the way in which this social need determines initial conceptualizations. Consider our initial conceptualization of WOLF. There was presumably a large field of unconscious perception that included the wolf, some trees, some plants, and so on. We did not conceptualize at once every detail of this field of unconscious perception, nor did we get anywhere near doing so. We simply conceptualized one articulated object that was before us that we had a *need* to communicate to others (and we conceptualized it in a rapid, "good enough" way). There was no initial motivation to conceptualize from our unconscious perception anything *but* what was needed for this "social or herd utility" (GS 354).

A point that follows from what I have said above, but which I have not explicitly brought out, is that this "need to communicate" is what drove all initial consciousness and thus what drove *self-consciousness* as well. We were compelled to form initial conceptualizations of ourselves based on our need to communicate certain states of ourselves to others. So, Nietzsche argues, "The human being inventing signs is at the same time the human being who becomes ever more keenly conscious of himself. It was only as a social animal that man acquired self-consciousness—which he is still in the process of doing, more and more" (GS 354). In other words, we conceptualized something about ourselves (i.e., created a word that we could employ in abstract thought) only insofar as we were put in a position of needing to communicate that aspect of our unconscious perception to someone else. Here it helps to remember that, for Nietzsche, we can, in some sense, be aware of our unconscious perceptions. So, initially, we are

aware of some unconscious perception of ourselves, but we do not think about ourselves using abstract thought, that is using concepts (which, for Nietzsche, means think about ourselves *consciously*). Eventually we need to communicate a certain state about ourselves, for example 'hunger', and so we have to *fix* this aspect of our unconscious perception of ourselves using a certain fixed sound and to conceptualize it as HUNGER, so that we can express this state to others. Nietzsche's point here is that, since this is the only way we created these fixed ideas, which are employable in thought, this is the only way we can consciously think of ourselves (since conscious thought, for Nietzsche, is thinking that uses concepts). What is additionally interesting about the above quote on "self-consciousness" is that Nietzsche claims this process of coming to self-knowledge is something humans are "still in the process of doing, more and more." It should be no surprise that Nietzsche thinks that our understanding of ourselves is still underdeveloped. What is surprising, however, is that Nietzsche seems to be indicating that there has been, and is, *progress*, in our coming to self-consciousness.

Nietzsche shows what follows from the preceding point—that our self-consciousness is initially determined by social need—in a dense section that recalls the title of GS 354, "*On the 'genius of the species'"*:

> Consequently, given the best will in the world to understand ourselves as individually as possible, "to know ourselves," each of us will always succeed in becoming conscious only of what is not individual but "average." Our thoughts themselves are continually governed by the character of consciousness—by the "genius of the species" that commands it—and translated back into the perspective of the herd. (GS 354)

Here Nietzsche is concerned with driving home an important point about our initial conceptualizations of ourselves. If conceptualization is motivated by the need to express states of ourselves to others, we will quickly and roughly conceptualize one aspect of our unconscious perception of ourselves. We communicate FEAR, when our actual unconscious perception before the conceptualization process is much more complex. In other words, there is no motivation (initially at

least) to conceptualize an aspect of ourselves unless it is essential to herd utility. If there is a complexity to our fear, it is lost by the general, rough, communicable term 'fear'. Furthermore, if there is an aspect of ourselves that is so unique that it is incommunicable, or, as is more often the case, we have no social reason to communicate it, this aspect of ourselves remains "un-known" to us consciously.

We have additional concern about the state of our self-consciousness if we consider the social way our concepts are refined. I mentioned before that, perhaps, when we use our concepts with others this can, over time, cause our conceptualizations to become more accurate. So, as we use our concept TREE and others use their concept TREE, we come closer to an accurate conceptualization of TREE. With this TREE example it seems we have combined some of the intricacies of our individual conceptualizations and eliminated the aspects of that concept that are non-essential or that do not match up with others' individual conceptualizations. But, it can just as well have the opposite effect, especially when dealing with the internal, individual world. While trees are external to all humans and each human can in principle have access to the same exemplars of the concept, I am the only one who could know what it feels like to "be me. When we consciously converge on concepts about ourselves, for instance, by expressing and attempting to relate our FEAR with another person's FEAR, we end up with concepts that are not individual, but average. We start to understand FEAR *not* as related to the unconscious perception that was unique to each of us, but rather, our concept FEAR takes on a meaning that derives from the similarities among our experiences. Now, when we attempt to think about ourselves using the concept FEAR, we are self-conscious only of something that is "average," that is, of the aspects of FEAR that everyone's individualized conceptualizations share.[22]

[22] Here, again, Nietzsche's early remarks in "On Truth and Lie in the Extramoral Sense" are helpful. He claims that with a concept like HONESTY, we have forgotten the individual differences in each experience we fixed as HO-

This understanding of how we become self-conscious leads to Nietzsche's next claim in GS 354 that, "Fundamentally, all our actions are altogether incomparably personal, unique, and infinitely individual; there is no doubt of that. But as soon as we translate them into consciousness *they no longer seem to be*." Given my interpretation above, it is not hard to see what Nietzsche is talking about. Our unconscious perception of ourselves, how we actually are, is unique and "incomparably personal"; we become conscious of ourselves only insofar as it is required for social utility and herd purposes and only via concepts that relate necessarily to what is comparable and common. Initially, at least, we have no motivation or reason to perceive consciously, that is to form concepts of, what is unique in our actions. Consciousness, and this early conceptualization of our selves, is not aimed at conceptually understanding our true, unique selves.

We are now in a position to look at, and make sense of, the title of GS 354—"*On the 'genius of the species'."* "Our thoughts" Nietzsche claims, "are continually governed by the character of consciousness—by the 'genius of the species' that commands it—and translated back into the perspective of the herd." By "genius of the species," Nietzsche is not, of course, referring to a conscious individual genius, but rather the social needs that require or "command" certain unconscious perceptions to be conceptualized.[23] So in this sentence, which includes the title of the section, both of the points I have made about the ways conceptualization can go wrong are brought out. First, the character of consciousness or "genius of the species," determines *what*

NESTY. This leads us to the illusion that we know some "*qualitas occulta*," when, in fact, we "know only numerous individualized, and thus unequal actions, which we equate by omitting the unequal and by then calling them honest actions" (TL).

[23] Nietzsche cannot, of course, claim that this is true of all thoughts, since some thoughts can occur unconsciously, as he says in the beginning of this section. This is an example of Nietzsche not being as careful as we would like with his use of words (such as 'consciousness', 'will', and 'freedom'). When Nietzsche uses "thoughts" here we can think of conceptualized thought—that is, thought that we can express in the form of words and employ in abstract thought. When Nietzsche talks about being able to "think" unconsciously, he is referring to thinking that does not happen with words and cannot be used in abstract thought.

content is going to be conceptualized. Secondly, *how* this content is conceptualized is determined by the "perspective of the herd," that is by social need and *not* by a motivation for accuracy.

3. The Conceptualization of Bad Conscience

I have so far shown how consciousness was initially developed and outlined the problems this development causes for the accuracy of our concepts. Far from being epiphenomenal, consciousness plays a crucial role in Nietzsche's genealogy of morality, and, I will argue, in his overall goal of a revaluation of all values. By showing briefly how the above theory can connect Nietzsche's disparate comments about consciousness with a central aspect of his philosophy as a whole, I will have given further support to a reading of Nietzsche according to which he has a novel, naturalistic, causally efficacious view of consciousness that it is important not to overlook.

We saw in section III.1 how the origin of the "bad conscience," a central idea in Nietzsche's genealogy, was due to the internalization of the aggressive instincts—an internalization caused by the change of man from a "beast of prey" to a social creature (GM II 16). This change was the catalyst for the expansion of an entire inner world. Now, as Nietzsche claims, humans were "reduced to their 'consciousness', their weakest and most fallible organ!" (GM II 16) With the theory of consciousness that I laid out above, we now know why Nietzsche considers consciousness weak and fallible: the content of consciousness, our concepts, are made up of only part of our unconscious perception and they capture the world only in whatever way is needed for societal communication. The bad conscience, however, creates *needs* for concepts that are more complex than the vague "social need" described above. In what follows, I will explain how the bad conscience works and the *need* it has for concepts.

As I have already discussed, for Nietzsche, instincts cannot be extirpated, only redirected. When society forced the aggressive instincts inside, man suffered "of himself," and this *feeling* is what is called bad conscience. Katsafanas sums up nicely everything that goes into this feeling of bad conscience:

> The bad conscience is a medley of all of this: the pain engendered by the internalization of the aggressive instincts, the feeling of being turned against a part of oneself, the feeling of internal discord, the feeling of being a threat to oneself, and the feeling of being a threat to society. Fundamentally, then the bad conscience is a complex affect, engendered by the feeling of the instincts' being at odds with one another. (2005: 20)

As Katsafanas points out, it is only the aggressive instincts that are internalized, so the other instincts, such as those aided by, or identified with the social aims, are not internalized but rather remain and are at war with the aggressive instincts. The social instincts (more properly understood as instincts connected to social aims) that remain are what, *in the person*, prevent the aggressive instincts' being discharged outwardly. Katsafanas explains the paradoxical internalization that this causes:

> the social instincts come to include, as an essential component, a drive to repress the outwardly-directed aggressive instincts; and this new drive causes intense suffering. This new, pain-inducing drive is just the aggressive instincts themselves, in an internalized form. (20)

In other words, our social instincts *prevent* the aggressive instincts' outward expression, by giving the aggressive instincts an alternative outlet of expression: namely, condemning the urge for their own outward expression.

In the background here is Nietzsche's view of the self. When Nietzsche talks positively about the Self he always equates it with some sort of hierarchy of drives.[24] Unfortunately, he uses other key terms in very similar ways when he talks about the Self. In section 19 of *Beyond Good and Evil*, for instance, he refers to the body as a "social structure composed of many souls," and in section 12 of *Beyond Good and Evil* he offers a hypothesis about the soul accord-

[24] Katsafanas claims that we can consider drives and instincts as interchangeable in his discussion of Nietzsche's theory of consciousness, even though he recognizes there are important differences (2005: 29fn32).

ing to which the soul is a "social structure of the drives and affects." These two views I consider to be equivalent and importantly related to his notion of Self. The fact that he refers to the body as composed of souls and the soul as composed of drives does not mean he takes this to be a three-tier relationship; rather, Nietzsche refers to the Self using different terms, such as 'soul', and this Self is composed of many different parts, which he calls at various times 'instincts', 'drives', 'affects', 'wills', and 'souls'. There are indeed important, nuanced differences in how he uses all of these terms, but for our purposes, I will focus on only one important distinction: the distinction between a drive and an instinct.

An instinct is a deeply rooted desire to act in some general way and with a fixed, unalterable amount of force. An aggressive instinct, for example, is, the desire humans *necessarily* have to act aggressively. It is *general*, because the desire to act aggressively does not have a specific target. Together the instincts comprise, for Nietzsche, all impetus for action. So, when Nietzsche later begins to describe a "will to power," I take this actually to be an instinct, which Nietzsche presents as the ultimate instinct—the one instinct that can explain, at a deep level, how we act. A drive, on the other hand, is how an instinct finds expression, and a drive will always have a more specific target. So, the social instincts are more properly understood as drives—that is, as instincts that express themselves in the specific aim of pro-social behavior. This is an important distinction overlooked by Katsafanas. By equating drives and instincts, as he does, his account appears to lead to a contradiction. If an instinct, as he rightly claims, is something that "cannot be straightforwardly eliminated, but only restrained or redirected" (2005: 19), then it does not make sense for us to refer to some instincts as "social instincts" without further explanation, since that would mean those instincts did not exist before man became a social animal. A drive is contingent, an instinct is not. This point becomes clearer when we consider the drive

known as "the will to knowledge."[25] According to Nietzsche, humans have not always had a will to knowledge; the will to knowledge is something that can increase during one's life and it is found in varying degrees of strength in different people. It is not, therefore, something that has a determinate quantity of force. It follows from this that every instinct will be expressed in the form of a drive or drives and that every drive is connected to an instinct. With these distinctions in mind, we can return to our discussion about the "internalization of man" which is the origin of the bad conscience.[26]

Using these new terms, the "bad conscience" can be understood as the unconscious perception that centers around a conflict between the social drives and the outward directed aggressive drives. The social drives, in effect, win out. This cannot, of course, mean that there was a change in the *force* of the underlying instincts; rather, the social drives "win out" by getting[27] the expression of the aggressive instincts to change from outward-directed aggressive drives, to inward-directed aggressive drives.

The conceptualization of the unconscious perception of bad conscience is tricky. To get a grasp on this we need to look at sections immediately preceding GM II 16 and see what Nietzsche says about the conceptualization of punishment in general. In GM II 12, Nietzsche first distinguishes between the origin and the purpose of punishment. Contrary to what is commonly thought of as the relationship between origin and purpose, Nietzsche declares:

---

[25] I have not introduced 'will' as one of these technical terms, since Nietzsche uses it in *several* different ways and it would be needlessly confusing. Most of the time he uses "will to...," such as in "will to knowledge" or "will to nothingness." In these instances, he is referring to what I have just described as a drive; however, even here Nietzsche is not ideally consistent, as the "will to power" example shows.

[26] The two-part distinction I have made here does not entail that there are not other distinctions within the categories of "drives" and "instincts." Certainly, with drives, there are hierarchies, and as I already mentioned with reference to "the will to power," some instincts are "deeper" than others.

[27] Of course, *how* the social drives "get" the aggressive instincts to alter the direction of their expression is a separate issue that is not important for this paper's thesis. One way to make sense of it, however, is to say that the social drives are not, in fact, *causing* this change, but rather, a brutal ruling class *causes* this change by brutally punishing the outward-directed aggressive drives and creating drives that are subservient to the social drives.

> the cause of the origin of a thing and its eventual utility, its actual employment and place in a system of purposes, lie worlds apart; whatever exists, having somehow come into being, is again and again reinterpreted to new ends, taken over, transformed and redirected by some power superior to it. (GM II 12)

In his talk about something's being "reinterpreted," Nietzsche is referring to our concepts, and with the model described in the previous section, we can understand new interpretations as new conceptualizations. This reading is supported with the next paragraph:

> purposes and utilities are only *signs* that a will to power has become master of something less powerful and imposed upon it the character of function; and the entire history of a "thing," an organ, a custom can in this way be a continuous sign-chain of ever new interpretations and adaptations. (GM II 12)

A couple of important points can be made in regard to this passage. First is Nietzsche's use of and emphasis on the word "sign." Nietzsche's references to consciousness consistently refer to it as the "sign world." We saw this in GS 354: "conscious thinking takes the...*signs of communication*," the human being becoming conscious is "the human being inventing signs," and "the world of which we can become conscious of is only a surface and sign-world." So the concepts with which we understand purposes and utilities are not the ultimate cause of the action (punishment in this case), but rather are the result of conceptualizations of the unconscious perception of punishment. *How* this unconscious perception is conceptualized is determined by the ruling drive. In the case of the unconscious perception of bad conscience, it is the social drives that conceptualize the unconscious perception as a feeling of guilt, since this helps fulfill the social drives' goal.

This brings up an important point about the conceptualization process that was not brought up in the analysis of GS 354. While social drives are what drove *initial* conceptualizations, these are not the only things that can drive conceptualizations. Out of this whole developmental process, other drives, such as a drive to knowledge, play the role of "ruling drive" in one's hierarchy and can therefore determine how an unconscious perception is conceptualized.

In GM II 13 Nietzsche repeats a notion that he first brought up in *The Wanderer and His Shadow*, that "all concepts in which an entire process is semiotically concentrated elude definition; only that which has no history is definable" (GM II 13; WS 33). This is the same idea he alludes to in the second part of the passage from GM II 12 quoted above. The history of a concept, such as "punishment," can be a "continuous sign-chain of ever new interpretations" (GM II 12). This chain makes that concept indefinable. The concept has no precise meaning that can be given, since it has been subject to so many different interpretations. In GM II 13, Nietzsche sets out to show this point by giving eleven interpretations or "meanings" that punishment has taken on in the history of its use as a concept and states at the beginning of GM II 14 that "this list is certainly not complete." The points to note for this paper's purpose are first, that Nietzsche is giving one more reason to be skeptical of the accuracy of our conscious mental states: our concepts (which, for Nietzsche, make up the entirety of conscious mental states) carry historical baggage and each concept is only one of many possible interpretations of a certain unconscious perception. The second point to draw from Nietzsche's discussion of punishment here is that we should in no way expect to be able satisfactorily to trace the contents of our consciousness (our concepts) from the origin described above, in simple social utility, to the complex and indefinable nexus of concepts we use today.

In GS 354, we saw that what drove our conceptualization was not a desire for accuracy or completeness, but what we were generally referring to as social need. As I have just demonstrated, the drives that determine how and what we conceptualize are, in fact, much more complex than just the broad term "social need" suggests. Although the general idea—that what will allow for the survival of the species is what is conceptualized (a point I supported in section III.2)—is

correct with respect to our initial conceptualizations[28], after these conceptualizations, other drives that are not aimed directly at species survival develop and play a role in how the conceptualization process works. These various drives significantly complicate any attempt to figure out what drives are at work in any one particular conceptualization. This is what Nietzsche's example of the many meanings given to punishment shows.

The conceptualization of punishment as GUILT is what Nietzsche is leading up to in the sections before GM II 16. In GM II 14, after laying out all the meanings that have been given to the act of punishment, he explains that there is one "*supposed* utility" which punishment "always finds its strongest support in." Nietzsche describes it this way:

> Punishment is supposed to possess the value of awakening the *feeling of guilt* in the guilty person; one seeks in it the actual *instrumentum* of that psychical reaction called "bad conscience."[29] (GM II 14)

So punishment, Nietzsche claims, is predominantly used to awaken the feeling of guilt, which Nietzsche relates to the growth of the "bad conscience" as a whole.

The conclusion to all of this is that the unconscious perception of punishment, inherent to the development of the "bad conscience," is, in our conceptual framework, conceptualized predominantly as GUILT. Historically, at least, we take punishment predominantly to mean, to *signify*, the attribution of guilt. This has profound effects according to Nietzsche. At the end of GM II 14, he describes how punishment worked in "the millennia *before* the history of man," when it was not conceptualized as GUILT. "The 'bad conscience,'" Nietzsche says, "this most uncanny and most interesting plant of all our earthly vegetation, did *not* grow on this soil" (GM II 14),

---

[28] By initial conceptualizations, I am referring to the very first conceptualizations. These were conceptualizations like HUNGER and WOLF, and the social need that necessitated them is what can make sense of consciousness developing at all (and is what Nietzsche refers to in GM II 16, when he claims that man was forced to rely on his consciousness).

[29] For Nietzsche, the idea of desert is inextricably tied to this idea. In GM II 4, he claims that when punishment is understood as guilt, there appears on earth this idea that "'the criminal deserves punishment *because* he could have acted differently'."

rather, he is implying, the bad conscience grew when the unconscious perception of the bad conscience was conceptualized as guilt.

Explaining how and why it was conceptualized as guilt is the main task of GM II. I have already given a brief overview of this above: the social drives conceptualize the unconscious perception as GUILT, since this helps the social drives reach their aim. While it would go beyond the purposes of this paper to capture this entire development and all of the details Nietzsche discusses, we can get an idea of what Nietzsche has in mind by picking up this story in section GM II 22.

In the sections leading up to GM II 22 Nietzsche presents a macro-level explanation of guilt arising as an increasing feeling of indebtedness. In the first stage this is a feeling of indebtedness to one's tribal ancestors as those to whom one's tribe owes its survival and success, but in its final form this becomes the ultimate feeling of indebtedness: debt before the Christian God. Then, in GM II 22, Nietzsche presents a micro-level explanation, or, in his words, an explanation of what "really happened here, *beneath* all this." Nietzsche explains this in what can be considered the climax and partial summation of book II, which deserves to be quoted at length:

> the creature imprisoned in the 'state' so as to be tamed, who invented the bad conscience in order to hurt himself after the *more natural* vent for this desire to hurt had been blocked—this man of the bad conscience has seized upon the presupposition of religion so as to drive his self-torture to its most gruesome pitch of severity and rigor. Guilt before *God*: this thought becomes an instrument of torture to him. He apprehends in "God" the ultimate antithesis of his own ineluctable animal instincts; he reinterprets these animal instincts themselves as a form of guilt before God. (GM II 22)

Nietzsche continues, "In this psychical cruelty there resides a madness of the will" that "infect[s] and poison[s]...the fundamental ground of things with the problem of punishment and guilt so as to cut off once and for all his own exit from this labyrinth of 'fixed ideas'." Nietzsche concludes the paragraph with the following pithy summation, "what *bestiality of thought* erupts as soon as he is prevented just a little from being a *beast in deed!*" (GM II 22)

What Nietzsche is claiming here is that when the outward-directed aggressive drives are "blocked" (GM II 22),[30] the aggressive instincts that initially found expression in these drives found expression in inward-directed aggressive drives, which were aimed at hurting oneself. The most effective way these drives could work to express the aggressive instincts was by conceptualizing the unconscious perception of pain caused in the "blocking" of the outward-directed aggressive drives as one's *own* fault. So this unconscious perception of a brutal ruling class punishing outward-directed aggressive behavior could have been conceptualized in different ways and, Nietzsche wants to say, *was conceptualized* differently at times. The predominant conceptualization, however, was the conceptualization of it as GUILT—and eventually as the most extreme guilt: guilt before God.[31]

Two important questions remain about the account just given, and in answering them, I will show how this story connects to one of Nietzsche's main insights and the important role conscious mental states play in that insight. The first question is: "Exactly *why* are we driven to the conceptualization of guilt before God?" The second question is: "Exactly *how* are we driven to this conceptualization?" Before answering the first question, I will summarize what I have already shown about this process. The bad conscience is the aggressive instincts turning back against themselves. This "turning back" is more properly understood as the consequence of the aggressive instincts' expression in outward-directed aggressive drives being blocked, an event that in turn caused a new expression of the aggressive instincts in inward-directed aggressive drives that find the subject guilty for having the outward-directed aggressive drives. The reason why the guilt is increased to extreme proportions has to do with the joy one feels in expressing

[30] As mentioned previously, Nietzsche offers some examples of how this "blocking" probably happened in GM II 3. Here, Nietzsche gives us graphic descriptions of "I will not's" being ingrained in man; these descriptions involve such things as the "flaying alive" of a criminal and boiling a criminal in oil.

[31] So, initially, the punishment was from external authority, a "ruling class," but eventually man punished himself.

their aggressive instincts, something Nietzsche calls "the enjoyment of violation" or "*de faire le mal pour le plaisir de le faire* [Of doing evil for the pleasure of doing it]." The inward-directed aggressive drives, as the new expression of these aggressive instincts, will continue to push for more and more guilt, since the more guilty one feels the more fully the aggressive instincts are expressed.[32]

I have already mentioned the short answer to *how* this ultimate feeling of guilt can be obtained, which is to consider oneself guilty before God. Much more, however, can be said, and is said by Nietzsche, about this process.

At first we feel guilty only about certain outward-directed aggressive drives. In order to increase our feeling of guilt, our inward-directed aggressive drives conceptualize unconscious perceptions in such a way as to make all of our natural instincts "bad." We subscribe to certain ideals because they are *not*-natural—that is, they are *not* what we and life are actually like. In short, instead of understanding concepts that are good as concepts relating to *this* earthly life of constant change, eventual death, some form of causal determinism, and so on, we consider concepts good that relate to life *beyond* this by positing "fixed ideas," eternal life, a *causa sui* ability and so on.

An important point for Nietzsche is that the conceptualization process is such that it encourages this view of the beyond. We form a concept, he claims, "through an arbitrary abstraction from...individual differences" (TL, p.46). Using our concept "leaf" as an example, Nietzsche explains what happens when we "forget" the differences between leaves:

---

[32] This is a really thorny area of *The Genealogy of Morals*, and one that raises many interesting questions in its own right. For my purposes it is not necessary to attempt to answer questions about how to make sense of one's "joy in cruelty" and the instincts' always desiring further expression; rather, for my paper, it is necessary to note only that this *is* indeed Nietzsche's view (GM II 5; GM II 6). One way to begin an answer, however, is to note how the will to power, Nietzsche's ultimate instinct, is related to joy. In *The Antichrist*, for example, he claims that happiness is "the feeling that power is *growing*" (A 2).

> it gives rise to the idea that in nature there might be something besides the leaves which would be "leaf"—some kind of original form after which all leaves have been woven, marked, copied, colored, curled, and painted, but by unskilled hands, so that no copy turned out to be a correct reliable and faithful image of the original form. (ibid.)

In other words, our ontology increases to include a world of "forms," when, in reality, no *things* correspond to such pure concepts. The ultimate "form," of course, is God. God symbolizes everything we are not and in forming the concept of God, our inward-directed aggressive drives have found the fullest way in which they can express the aggressive instincts.

The role of consciousness in this process is undeniable. To review: for Nietzsche, what consciousness *is* is a mental state with the property of being conceptually articulated. Thus, when Nietzsche discusses the formation of concepts and that concepts (such as punishment) can be conceptualized in different ways ("interpreted" is the term Nietzsche uses), he is talking about conscious mental states. If we jump back to the first section of the second essay we can see clearly in his introductory remarks on the "memory of the will" (GM II 1) where Nietzsche places this new inner world, including consciousness, in relation to action:

> ...between the original "I will," "I shall do this" and the actual discharge of the will, its *act*, a world of strange new things, circumstances, even acts of will may be interposed without breaking this long chain of will.

This section, and especially the section after it, which discusses the "right to make promises" (GM II 16), is an oft-discussed part of the *Genealogy* and there are many radically different interpretations in the literature. While I by no means feel I have settled the debate, with the understanding of consciousness and the entire inner world that I have put forth, I have presented a very plausible way to begin making sense of this "world of strange new things" (GM II 1).

One thing that Brian Leiter claims and that my analysis makes clear is that consciousness is not *by itself* causally efficacious. However, the view of the "will as secondary cause" that I put forth at the beginning of the paper fits with what he says in the passage from *The Genealogy*

of Morals just quoted. Here, Nietzsche's claim is that there is an original "I will," but instead of this "I will" going straight through to action, as it had always done in our "animal past" (GM II 16), this "world of strange new things" has been interposed. But viewing "the will as a secondary cause" can also be misleading. There is no separate faculty called "consciousness" that receives concepts and processes them using consciousness-specific tools such as reason and thinking. According to Nietzsche, we can in some sense be aware, reason, and think, *without* consciousness. Conscious thought and conscious reasoning are just thought and reasoning that happen to make use of concepts. Further, these conscious mental states come about only via conceptualizations that are directed by one's unconscious drives. Nietzsche makes this clear in a typical passage in which he criticizes the view that consciousness is a separate faculty: "a directing committee on which the various chiefs of desires make their votes and power felt" determines our conscious content, and although often "one takes consciousness itself as the general sensorium and supreme court...it is not the directing agent, but an organ of the directing agent" (WP 524). The fact that our conscious mental states are determined by these other factors should neither surprise nor concern us. Our conscious mental states, like every other part of our body (and every other part of nature), are part of an interconnected, causally determined, physical system. What would matter, and make our conscious mental states epiphenomenal—at least in one sense of the word—would be our having conscious mental states that had no effect on the rest of the natural world;—in other words, consciousness would be epiphenomenal just in case nothing would change if we did not have these conscious mental states. My analysis shows that this is clearly not the case. I have shown, for instance, that if we interpret, that is conceptualize, an unconscious perception of punishment as GUILT, we not only feel the pain of the act of punishment, but, in addition, we feel the pain of being responsible.

Guilt, of course, is just an example. All of the conceptualizations of our unconscious perception are going to help determine our perspective on the world, how we communicate with one another, and, in some sense, how we act. While I have by no means explained every detail of the way concepts interact with the rest of one's body, it is clear that concepts *do* interact and play a role.

I will close this section by making some general remarks about the role these conscious mental states play. To reiterate a point from earlier in this paper, Nietzsche realizes that this new world of concepts—this "labyrinth of 'fixed ideas'" (GM II 23)—played an important role in our change from an individual "beast of prey" (GM I 11) into a social animal. In order to survive, we needed concepts, because we needed to be able to communicate. The inner world of which these concepts were a part, and which expanded as the need to communicate expanded, were part of "the most fundamental change [humans] ever experienced" (GM II 16). Often this creation of concepts, and the concepts that came out of this conceptualization, are viewed by Nietzsche as something dangerous and harmful. The main example is the account I gave above: the inward-directed aggressive drives created concepts of value that were *anti*-life (in order to be more cruel, by making one feel guilty for how they naturally are). However, Nietzsche does indicate that there may be a positive role for concepts as well. Nietzsche never elaborates on the details of this role, but it centered around what he called "a revaluation of values." It is clear that Nietzsche does not think that this is simply a matter of our deliberating consciously about value; anyone is capable of that. Rather, it will take someone with the right hierarchy of unconscious drives. It is also clear, however, that since we are "cut off once and for all" from an exit out of this world of concepts, this "world of strange new things" (GM II 1) is going to play an important role in the revaluation.

IV. CONCLUSION

In this paper I have shown how Leiter's epiphenomenalist reading of Nietzsche is not decisively supported by the text and the contemporary empirical evidence he cites. Nietzsche is committed to conscious mental states having causal influence. The text that Leiter cites and many other of Nietzsche's negative remarks about consciousness are directed at a view of consciousness as a separate faculty that is cut off from the rest of the body and that is self-caused. In other texts, including in several key passages in *The Genealogy of Morals* and *The Gay Science*, Nietzsche offers a naturalistic, evolutionary account of how our conscious mental states have developed and the function they perform in relation to action. In these passages Nietzsche refers to consciousness not as a separate faculty, but rather as the property of being conceptually articulated that some mental states have. The theory I have outlined above is in no way complete, and there is much promising work to be done filling in many of the details of the theory and in extending this view to other aspects of Nietzsche's thought. What this paper has sought to show is that in a very important way, the epiphenomenalist reading of Nietzsche misses the mark. It ignores Nietzsche's thoughtful, novel remarks regarding the nature, development, and influence of conscious mental states. By presenting a theory of consciousness that incorporates these insights, I have set the stage for a more complete and powerful interpretation of Nietzschean action.

V. REFERENCES

1. Nietzsche's Texts

I have followed the following abbreviations when referring to Nietzsche's texts:

A	*The Antichrist*
BGE	*Beyond Good and Evil*
D	*Daybreak*
GM	*On the Genealogy of Morals*
GS	*The Gay Science*
TI	*Twilight of the Idols*
TL	*On Truth and Lie in the Extramoral Sense*
WP	*The Will to Power*
Z	*Zarathustra*

I have used the following translations of Nietzsche's texts:

The Antichrist, in *The Portable Nietzsche* (1954).

The Basic Writings of Nietzsche, edited and translated by W. Kaufmann, New York: Modern Library, 1967.

Beyond Good and Evil, in *The Basic Writings of Nietzsche* (1967).

Daybreak, translated by R.J. Hollingdale, edited by M. Clark and B. Leiter, Cambridge: Cambridge University Press, 1997.

On the Genealogy of Morals, in *The Basic Writings of Nietzsche* (1967).

The Gay Science, translated by W. Kaufmann, New York: Vintage, 1974.

The Portable Nietzsche, edited and translated by W. Kaufmann, New York: Viking, 1954.

The Twilight of the Idols, in *The Portable Nietzsche* (1954).

On Truth and Lie in the Extra Moral Sense, in *The Portable Nietzsche* (1954).

The Will to Power, translated by W. Kaufmann and R.J. Hollingdale, New York: Vintage, 1968.

Thus Spoke Zarathustra, in *The Portable Nietzsche* (1954).

2. Other Works Cited

Clark, M. and Dudrick, D. 2009. Nietzsche on the Will: An Analysis of BGE 19. In K. Gemes & S. May, *Nietzsche on Freedom and Autonomy* (pp. 247-268). New York: Oxford University Press.

Graham, G. 1993. *Philosophy of Mind: An Introduction*. Cambridge, MA: Blackwell Publishers.

Katsafanas, P. 2005. Nietzsche's Theory of Mind: Consciousness and Conceptualization. *European Journal of Philosophy, 13* (1), 1-31.

Leiter, B. 2002. *Nietzsche on Morality.* New York: Routledge.

Leiter, B. 2009. Nietzsche's Theory of the Will. In K. Gemes, & S. May, *Nietzsche on Freedom and Autonomy* (pp. 107-126). New York: Oxford University Press.

Libet, B. 1999. Do We Have Free Will? *Journal of Consciousness Studies , 6* (8-9), 47-57.

Mele, A. 2010. Scientific Skepticism About Free Will. In T. Nadelhoffer, E. Nahmias, & S. Nichols (Eds.), *Moral Psychology: Historical and Contemporary Readings.* 1-14: Wiley-Blackwell.

Nahmias, E. 2010. Scientific Challenges to Free Will. In T. O'Connor, & C. Sandis (Eds.), *A Companion to the Philosophy of Action* (pp. 345-356). Wiley-Blackwell.

Strawson, G. 1998. Luck Swallows Everything. *TLS* , 8-10.

Wegner, D. M. 2008. Self is Magic. In J. Baer, J. C. Kaufman, & R. F. Baumeister (Eds.), *Are We Free?: Psychology and Free Will* (pp. 227-247). Oxford: Oxford University Press.