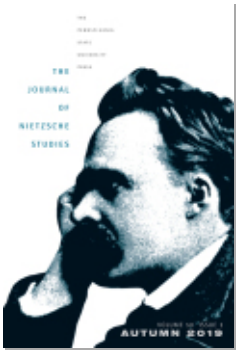
Nietzsche on the Origin of Conscience and Obligation

Avery Snelson



The Journal of Nietzsche Studies, Volume 50, Issue 2, Autumn 2019, pp. 310-331 (Article)

Published by Penn State University Press

Nietzsche on the Origin of Conscience and Obligation

AVERY SNELSON | UNIVERSITY OF CALIFORNIA, RIVERSIDE

Abstract: The second essay of Nietzsche's *Genealogy of Morality* (*GM*) offers a naturalistic and developmental account of the emergence of conscience, a faculty uniquely responsive to remembering and honoring obligations. This article attempts to solve an interpretive puzzle that is invited by the second essay's explanation of nonmoral obligation, prior to the capacity to feel guilt. Ostensibly, Nietzsche argues that the conscience and our concept of obligation originated within contractual ("creditor-debtor") relations, when creditors punished delinquent debtors (*GM* II:5). However, this interpretation, which I call the contractualist reading, is incoherent and subject to an insoluble bootstrapping problem. I argue instead that Nietzsche provides two accounts of nonmoral obligation in the second essay, and that the conscience originated in the morality of custom to track rule prohibitions ("I will nots" [*GM* II:3]), which Nietzsche conceives of as involuntary or reciprocal obligations that, unlike contractual debts, do not require the making of promises.

Keywords: conscience, obligation, contractualist reading, rules, debts

The second essay of *GM* is an ambitious text structured around explaining the origins of guilt, though such an account is not actually offered until the reader is roughly 80 percent of the way through the story. Prior to that point the essay provides a history of human punishment and socialization and profiles the emergence of conscience. Like the essay as a whole, Nietzsche's developmental account of the conscience is fragmentary and immensely complex; it takes on numerous forms and goes through various stages of development throughout the essay, in response to different environmental pressures. My focus here is on the most incipient, nonmoral form of conscience Nietzsche discusses, which predates the ability to feel guilt, the development of bad conscience, and the memory of the will.

DOI: 10.5325/jnietstud.50.2.0310
JOURNAL OF NIETZSCHE STUDIES, Vol. 50, No. 2, 2019

That Nietzsche is committed to such a conception is clear from a close reading of the first four aphorisms. The essay begins by foreshadowing the conscience as a "memory of the will" (*GM* II:1),[1] described as an ability to extend practical commitment in the absence of external incentives.[2] This form of conscience underwrites "permitted promising" (i.e., is the basis of trust) and belongs to the sovereign individual, the "'free' human being, the possessor of a long, unbreakable will, [who] has in this possession his *standard of value* as well" (*GM* II:2). The conscience did not *originate* as the "memory of the will," however. The sovereign's conscience is claimed to have been the product of a "long history and metamorphosis," originating as a memory of customary rule prohibitions ("I will nots"), understood by Nietzsche to be "primitive requirements of social co-existence" (*GM* II:3). In the opening line of the fourth aphorism, Nietzsche contrasts this memory of rules with the "bad conscience," the "consciousness of guilt" (*GM* II:4). Notably, at this point in the essay he then turns his attention for the first time toward the idea of debt and contractual relationships, arguing that both are precursors to the development of guilt (*GM* II:4, 6, 8).[3]

My aim in this article is to reconstruct the conditions in which this earliest form of conscience—the memory of "I will nots" (*GM* II:3)—would have originated, since the standard reading of the second essay's account of the emergence of conscience seems to me deeply flawed. In its broadest capacity the conscience is understood by Nietzsche to be a "consciousness of" one's obligations, or a kind of "memory" that takes as its object two distinct forms of obligation: rules and debts (*GM* II:3, 5). We acquired this memory, moreover, by being punished, but under what conditions were we punished? That is, under what conditions did the conscience and our concept of obligation originate? Ostensibly, Nietzsche thinks the conscience originated within the context of contractual relationships, or what he calls "creditor-debtor" (*GM* II:5) relationships, to ensure that we keep our promises to one another. I call this the *contractualist reading*.

I believe this interpretation is implicit in two recent articles by Bernard Reginster,[4] who has, in my opinion, offered the most thorough and compelling analysis of the memory of the will in the literature. Though I do not deny that the contractualist reading has a clear basis in Nietzsche's text, I argue in the second section that it is incoherent and subject to an insoluble bootstrapping problem: it requires that the debtor be able to communicate promises as a condition of forming contractual relationships, and contractual relationships are moreover necessary to explain the ability

to communicate promises. Accordingly, in the remainder of the article I develop an alternative interpretation to the contractualist reading.

I develop this alternative through a close analysis of the third aphorism, where Nietzsche appeals to the conscience to explain how humans became reliable, or "regular" and "predictable" (*GM* II:1, 2) in their behavior, which he claims is a "presupposition" to the memory of the will and the ability to make promises. We became reliable on his account by learning to conform our behavior to rule prohibitions, "I will nots" (*GM* II:3). In the third section, I explain this by showing that Nietzsche conceives of the "I will nots" as a species of what Frans de Waal calls "prescriptive rules," which are rules imposed by agents on other agents within dominance hierarchies, and that Nietzsche thus conceives of the conscience at this early stage in our moral development as conferring only an ability to remain conscious of social expectations more generally, and to conform our behavior to such expectations. Importantly, as I argue in the fourth section, these "I will nots" are conceived by Nietzsche as *involuntary* or *reciprocal* obligations that do not require the making of promises. Thus, I aim to show that Nietzsche offers a genealogy of *two* forms of nonmoral obligation in the second essay—involuntary rules and contractual debts—and since he takes the former to predate the latter, he is able to avoid the bootstrapping problem that plagues the contractualist reading.

The Contractualist Reading

According to the contractualist reading, the conscience was "bred" in humans unwittingly in response to the need to be able to keep promises, subsequent to forming contractual or "creditor-debtor" relationships (*GM* II:5). As Bernard Reginster has noted, "contractual relationships are established by *promising*, and so they involve the whole apparatus designed to make promising possible, particularly the recourse to the infliction of pain."[5] Here Reginster is using "promising" in two different senses. In the first instance he has in mind the act of *communicating* a promise, and in the second the ability to *keep* or sustain the motivation to fulfill one's promise. This latter ability, on his account, is what the conscience or "memory of the will" enables. "The 'will' to be remembered here," he says, "is an obligation undertaken, or a 'promise' made—'I will,' 'I shall do.'"[6] "The capacities that underwrite the right to make promises," he elaborates, "are not parts of the innate 'animal' endowment of human beings. They require that a new psychological structure be 'bred' into them, namely a 'conscience.' Conscience is a particular kind

of memory, which Nietzsche calls 'the will's memory.' . . . The purpose of the necessary breeding is thus to ensure not just the memory that a promise was made, but also the persistence of the motivation to keep it."[7]

Contractual relationships are instrumental to the development of conscience, according this view, because they are created when one person, the debtor, makes a promise to repay another, the creditor, for some good or service. Subsequent to making this promise, the creditor then held the debtor to that expectation of repayment, but since the debtor lacked a conscience, and therefore a "memory" of his obligation, he failed to repay and was punished by the creditor as an alternative means of compensation. This punishment, Nietzsche insists, was not intended to reform the debtor, teach him a lesson, or elicit feelings of guilt (*GM* II:14). It was administered simply because it was pleasurable, because it gratified the creditor's "instinct for cruelty" (*GM* II:5), and therefore provided him with an alternative form of repayment. However, because punishment was also a "mnemo-technique" (*GM* II:3), a procedure for creating memory, the creditor's punishment had the unintended effect of producing in the debtor a conscience, a faculty that allowed him to both "remember" his debt and sustain the motivation to keep his promise.

In fairness to Reginster, it must be said that he never explicitly endorses this reading. However, as his remarks above indicate, it seems to be implicit in his account of the development of conscience. In his more recent article, he does attribute the initial development of conscience to the morality of custom; however, he nonetheless maintains that it emerged due to the need for "promise keeping," and so it would seem that contractual relationships are still operating in the background.[8] More importantly, Nietzsche appears to endorse the contractualist reading himself:

> Calling to mind these contract relationships admittedly awakens various kinds of suspicion and resistance toward the earlier humanity that created or permitted them [. . .]. Precisely here there are *promises* made; precisely here it is a matter of *making* a memory for the one who promises [. . .]. In order to instill trust in his promise of repayment, to provide a guarantee for the seriousness and the sacredness of his promise, to impress repayment on his conscience as a duty, as an obligation, the debtor—by virtue of a contract—pledges to the creditor in the case of non-payment something else that he "possesses," over which he still has power, for example his body or his wife or his freedom or even his life [. . .]. Above all, however, the creditor could subject the

> body of the debtor to all manner of ignominy and torture [. . .]. (*GM* II:5)

Nietzsche does indeed seem to argue here that the act of *making* promises, because doing so created a dynamic in which creditors punished debtors and produced the conscience, made it possible to *keep* promises, just as Reginster's remarks above indicate. However, a consequence of this argument is that debtors first had to make promises of repayment they did not fulfill, prior to their being punished, as a condition of forming contractual relationships. For this reason, as I will now show, Nietzsche's explanation of origin of conscience and obligation is incoherent and self-undermining: he conceives of the ability to communicate promises as a condition of forming contractual relationships, and contractual relationships are moreover needed by him to explain the ability to communicate promises.

Promising is not simply the act of communicating an intention. As Reginster says, "To make a promise is to commit to doing something at some appointed future time even if doing so has by then become contrary to my 'private desires and advantages' (*D* 9)."[9] Promising involves expressing a commitment that is in a key sense *binding*, and that is what the debtor does in the above context. By making a promise, repayment is conveyed by the debtor and understood by both him and the creditor as something *nonoptional*, as a kind of *requirement*. This is because promising, unlike merely communicating an intention, involves communicating an *obligation*. Indeed, to make a promise just is to "communicate an intention to undertake an obligation."[10] Thus, to communicate a promise, the debtor must *already* have a concept of obligation.

I will have more to say about obligation in the next section, but here it bears emphasizing that we need not build too much into the idea. In particular, it need not be the case that the debtor understands himself to have a *moral* obligation to repay his creditor, or that he would feel guilty if he failed to do so. It need only be the case that the debtor recognizes that he is in a minimal sense bound to repay him: that he "should" or "ought" to do so, as Reginster says above, quoting Nietzsche, regardless of his "private desires and advantages" (*D* 9). Obligations have this character because they are *social requirements*, actions that I am bound ("should" or "ought") to perform by others regardless of my personal desires.[11] To be slightly more technical, the mental state of obligation, its unique sense of "ought" or "should,"

is a consideration given deliberative priority in order to secure reliability of behavior,[12] which, when acted upon, has the social function of satisfying the expectations of those who have power or authority. Communicating an obligation, then, requires the ability to discriminate between those courses of action that are purely personal—"hypothetical imperatives," as Kant would characterize them—and those that are nonhypothetical.[13] That is, promising requires that I be capable of recognizing and conveying the course of action as making a claim on me even if I am not inclined to do it, as having a special kind of priority given its social significance, as something that is recognized by me as required or "obligatory" in the sense that its performance is not solely up to me.

Above, the debtor's promise has this nonhypothetical character because it is made within a social dynamic in which the creditor will hold him to an expectation of repayment *regardless* of whether he wants to do so at the appointed time, and it is further assumed by Nietzsche that the creditor has the authority or the power to hold the debtor accountable should he fail to do so. Consequently, the idea that the debtor could make a promise in this context without having a familiarity with obligations as social requirements is simply *incoherent*. Not only this, it seems the debtor must have a conscience as well. If the debtor had no conscience, a condition for the possibility of obligation—specifically, a condition for the possibility of being *aware* of or having a "consciousness of" obligation—has not been realized. In such a world, the debtor would lack a "memory," not just of *his* obligation or debt, but of the very *idea* of obligation. The contractualist reading, because it maintains that the conscience and our concept of obligation originated as *consequences* of promising, gives rise to the following causal dilemma. It requires the truth of both of the following propositions, but these cannot be held consistently:

1. (P1) contractual relationships are formed by communicating promises ("precisely here there are *promises* made" [*GM* II:5]), and
2. (P2) the conscience and our idea of obligation originated as debt within contractual relationships.

If the idea of obligation originates as that of debt, the debtor must be capable of entering into contractual relationships, according to P2. But in order to enter into contractual relationships, the debtor must have some idea of

obligation to communicate promises, according to P1. And since contractual relationships are formed by communicating intentions to undertake obligations of repayment (P1), and because the debtor has no concept of obligation prior to entering contractual relationships (P2), this means it would have been *impossible* to form contractual relationships. Consequently, Nietzsche's account of the origin of conscience and obligation is subject to an insoluble bootstrapping problem on the contractualist reading.

I do think Nietzsche has the resources to avoid this problem, though doing so will require rejecting either P1 or P2, and thus rejecting the contractualist reading. What Nietzsche describes in the fifth aphorism is actually a fairly complex transaction, one that is created by the explicit expression of voluntary obligations and presumes, among other things, a cultural background in which laws, property, and money already exist. Though few scholars have taken note of it, he even assumes the debtor already has a conscience.[14] For these reasons, we might think that what he describes in this passage is a later development of a more rudimentary practice of "proto-promising" that does not presume the debtor has a concept of obligation or a conscience. I believe this to be the case, but even so it is hard to see how Nietzsche could reject P1. For one thing, he simply takes it for granted that promising is a condition of forming contractual relationships ("precisely here there are promises made"). Second, Nietzsche believes contractual relationships involve the kind of transfer of power or rights often thought to coincide with promising. They involve a transfer of property (collateral) and rights, the "directive and right to cruelty" (*GM* II:5), and as such coincide with the appearance of "legal subjects" (*GM* II:4) and "the most rudimentary form of personal legal rights" (*GM* II:8). Despite their "rudimentary" nature, these are complex ideas the like of which cannot be found in the animal kingdom, ideas that serve only to reinforce Nietzsche's claim that creditor-debtor relationships represent "man's preeminence with respect to other creatures" (*GM* II:8). Finally, as I will argue in the fourth section, "proto-promising" just is the practice of reciprocity, and Nietzsche's account of reciprocal obligation is offered in the third aphorism in connection to the "I will nots." Consequently, I argue that Nietzsche rejects P2, which would require that he recognize the existence of nonmoral obligations that are more basic than, and indeed preliminary to, voluntary debts. In the next section, I aim to show that he provides such an account in the third aphorism.

The Origin of Conscience and Obligation

Early in the second essay Nietzsche observes that being reliable, or "regular" and "predictable" (*GM* II:1, 2) in one's behavior, is a presupposition of promising, of being able to "vouch for [oneself] *as future*" (*GM* II:1). This is because, as we have seen, a person who promises commits herself to some future action regardless of her "private desires and advantages" (*D* 9), which requires an ability to discriminate the obligatory from the nonobligatory, an "ought" that is merely prudential or hypothetical from an "ought" that is, as I said above, "nonhypothetical." The idea underlying these remarks is that agents who are "regular" and "predictable" in their behavior are so because they do what is *expected* of them, because they conform their behavior to compulsory norms or rules, rather than act on their strongest desire or whim of the moment. Nietzsche's account of the conscience in the third aphorism is offered to explain how humans became reliable in these ways. There he conceives of the conscience as a kind of *social memory* that made it possible to follow rules and live with others.

> "How does one make a memory for the human animal? How does one impress something on this partly dull, partly scattered momentary understanding, this forgetfulness in the flesh, so that it remains present?" . . . As one can imagine, the answers and means used to solve this age-old problem were not exactly delicate; there is perhaps nothing more terrible and more uncanny in all of man's prehistory than his *mnemo-technique*. "One burns something in so that it remains in one's memory: only what does not cease *to give pain* remains in one's memory"—that is a first principle from the most ancient (unfortunately also longest) psychology on earth. [. . .] The worse humanity was "at memory" the more terrible is the appearance of its practices; the harshness of penal codes in particular provides a measuring stick for the amount of effort it took to achieve victory over forgetfulness and to keep a few primitive requirements of social co-existence *present* for these slaves of momentary affect and desire. [. . .] With the help of such images and processes one finally retains in memory five, six, "I will nots," in connection with which one has given one's *promise* to live within the advantages of society [. . .]. (*GM* II:3)

I will address Nietzsche's conclusion that the "I will nots" are "connected" with a "promise to live within the advantages of society" in the next section.[15] Here I will focus solely on the formation and character of this "memory." My aim is to show that the "I will nots" are basic to cooperative sociality, and as such explain the origin of obligation.

As John Richardson has remarked, "This memory is simply the ability to 'remember' social rules or practices, to be bound by them."[16] The conscience here consists simply in being able to remember a handful of customary prohibitions, rules like "I will not steal," "I will not kill," and "I will not lie," which Nietzsche considers to be necessary for the maintenance of social life. As I have already mentioned, a close reading of the first four aphorisms reveals that he takes the conscience at this stage to be preliminary to both the bad conscience and the memory of the will. In fact, Nietzsche claims this memory-making technique belongs to the "longest" and "most ancient psychology on earth." He attributes this "enormous work" to the "morality of custom," claiming it was the "true work of man on himself for the longest part of the duration of the human race, his entire *prehistoric* work" (*GM* II:2).[17] These remarks suggest that the above passage is concerned with some amorphous stage in our evolutionary past, perhaps as far back as the appearance of *Homo*, what Nietzsche above calls the "human animal."[18] Speculation aside, that Nietzsche intends for this form of conscience to extend very far back in our evolutionary heritage cannot be questioned: the capacity that it is invoked to explain—the ability to follow rules and be reliable in one's behavior—is not distinctly human.[19]

As Frans de Waal remarks, "All animals conform to social rules. That is, their conduct toward conspecifics is to some degree predictable."[20] Rules, so understood, are regularities that circumscribe behavior,[21] and de Waal refers to rules imposed by agents on other agents as *prescriptive rules*, which is what Nietzsche describes above. De Waal moreover believes that prescriptive rules generate obligations, or possess an "ought quality," because they are learned behavioral patterns "actively upheld through reward and punishment." Such rules are made possible by hierarchical relationships in which A (typically a dominant) holds another B (typically a subordinate) to an expectation of conformity, which B (and others) learn by being punished.[22] (We will see evidence in a moment to suggest that Nietzsche also thinks the conscience developed within dominance hierarchies.) As de Waal concludes, "A prescriptive rule is born when members of the group

learn to recognize the contingencies between their behavior and that of [others] and act so as to minimize negative consequences."[23]

According to de Waal, obligations are generated whenever an agent with power or authority routinely holds another to an expectation of conformity, thereby constituting a compulsory norm or rule. H. L. A. Hart has defended a similar view of obligation. "Rules are conceived and spoken of as imposing obligations," he claims, "when the general demand for conformity is insistent and the social pressure brought to bear upon those who deviate or threaten to deviate is great."[24] Above, Nietzsche holds the same view regarding the "I will nots." These are conceived by him as "primitive requirements social co-existence," meaning that conformity with such rules is a prerequisite of community membership—they must be followed to live among others *at all*. Conformity with such rules is thus *obligatory* or *nonoptional*; indeed, a condition of cooperative sociality itself. At this level of analysis, obligation is understood to be a kind of *social fact* corresponding to a rule, an action that is required within a domain or by an activity merely in virtue of being a condition of its existence.

The conscience is invoked by Nietzsche in the third aphorism to explain how we became "regular" and "predictable" in our behavior, and so to explain how this strictly third-personal aspect of obligation became internalized and the human animal self-regulating. As Reginster correctly observes, this is essentially a matter of being able to "overcome" or "disregard" one's "private desires and advantages" (*D* 9),[25] but Nietzsche suggests we acquired this ability merely in virtue of the fact that rules must be followed to live with others, not due to the need to be able to keep promises. An "I will not" is understood by Nietzsche to be a *social requirement*, an action that I am bound ("should" or "ought") to perform by others regardless of my personal desires, the function of which is to secure reliable interactions with others as a necessary condition of cooperative sociality. So, obligation consists, in the first instance, in the *social fact* of obligation, in the fact that others with power or authority hold us to expectations, constituting rules.

However, obligation also consists, more interestingly, in a distinct *mental state*, an awareness of a course of action that "ought" to be done regardless of the agent's personal desires. Bernard Williams provides an accurate characterization of this "ought" by defining obligation as "a consideration given deliberative priority in order to secure reliability."[26] On his account, just as on Nietzsche's, the function of obligation is "to secure reliability, a state of affairs in which others can reasonably expect me to behave in

some ways and not in others."[27] Acting on an obligation, then, is not just an instinctual or automatic response; it requires the ability to reflect, to stand back, and to assess an action's social consequences. It requires, as Nietzsche says above, an ability to no longer be a "slave to momentary affect and desire" (*GM* II:3). As he claims, "With the help of this kind of memory one finally came 'to reason'!—Ah, reason, seriousness, mastery over the affects, this entire gloomy matter called reflection" (*GM* II:3).[28] By creating the possibility of deliberative conflict in the human animal, the development of conscience unified the third-personal and first-personal aspects of obligation in this way.

The conscience is an "inner voice" of our obligations, inculcated in us through punishment, but as I mentioned previously it is the voice of an "ought" that is not merely hypothetical. Nietzsche confirms this in a passage from *BGE*:

> Inasmuch as at all times, as long as there have been human beings, there have also been herds of men (clans, communities, tribes, peoples, states, churches) and always a great many people who obeyed, compared with the small number of those commanding—considering, then, that nothing has been exercised and cultivated better and longer among men so far than obedience—it may be fairly assumed that the need for it is now innate in the average man, as a kind of *formal conscience* that commands: "thou shalt unconditionally do something, unconditionally not do something else," in short, "thou shalt." (*BGE* 199)

In this quote, Nietzsche offers a description of the inner voice of conscience, a description of its *form*. He moreover tells us that the conscience originated to secure obedience to commands within social dominance hierarchies, which we have lived in since the inception of our species, and so the form this voice takes is that of a *command*: an "unconditional thou shalt."

What does this mean? First, we can be certain that Nietzsche does *not* mean the same thing as Kant. According to Kant, the categorical imperative is "unconditional" in the sense that compliance with it requires "the dissociation from all interest in willing from [from a motive of] duty."[29] For Nietzsche, this is a fiction—there is no "pure" moral motive. But he also recognizes, at least since *D*, that acting on an obligation is different from acting on the basis of hypothetical imperatives. There he appealed to the morality

of custom to offer a naturalistic explanation of the categorical force of moral norms, acknowledging that this cannot consist simply in considerations surrounding what is prudent or "useful" (*D* 9).[30] Hypothetical imperatives are generated *conditionally* on the basis of the agent's other desires or ends, and so if the agent's desires change or she gives up her end, the "ought" also disappears. As we saw in the case of the debtor's promise, obligations are not dependent on our desires in this way. So, what is essential to the idea of obligation and in need of a naturalistic explanation is this idea of an "ought" that is distinct from the merely prudential "ought," one that is in some sense not dependent on, nor entirely divorced from, the agent's other desires.

It is this gap that the conscience as a social memory is offered to fill in. The conscience is the inner voice of this nonhypothetical ought. It is nonhypothetical because, as Phillipa Foot observes, "Lacking a connection to the agent's desires or interests, 'should' in this case does not stand 'unsupported and in need of support'; it requires only the backing of the rule."[31] This does not imply that the imperative can be satisfied in the complete absence of the agent's other motives (e.g., fear and the drive for self-preservation), only that it does not *depend* on these for its existence. Also, as Nietzsche tells us above, conformity with a nonhypothetical ought requires regarding it not as a means to something else one wants, but as a *command*: a "higher authority which one obeys, not because it commands what is *useful* to us, but because it *commands*" (*D* 9). Importantly, obedience with a command need not be "pure." Commands may be obeyed out of fear, reverence, awe, or a mixture of these and other motives. What it means to obey a command "unconditionally" is simply that one complies with it in recognition of it *as such*, that is, by regarding it as something that "ought" to be done even when doing so is contrary to one's "private desires and advantages" (*D* 9). By doing so, one recognizes the existence of an "ought" that has a different status and significance than the "ought" of hypothetical imperatives.

Finally, I think Nietzsche offers a compelling explanation for how we came to act on the basis of such "oughts." He suggests that our doing so is simply a habit or tendency we have acquired to obey the commands of those who have rank, which he calls the "herd instinct of obedience" (*BGE* 199).[32] A command, being a compulsory order from a source of power or authority, when enforced consistently creates a norm or rule. And so the "herd instinct of obedience" is not merely a tendency to obey those with rank and to regard their commands as "unconditional," but a

tendency to conform one's behavior to norms or rules by regarding *them* as "unconditional." We became "regular" and "predictable" in our behavior, then, by developing this herd instinct of obedience. If this is right, the distinct "ought" of nonmoral obligation is simply the result this long history of "breeding" within social dominance hierarchies, a habit or tendency we have acquired to cognize some modes of conduct as *social requirements*. The nonhypothetical "ought" of obligation *just is* the voice of the "herd instinct of obedience."

Reciprocity and the Communal Bargain

I will now address Nietzsche's claim that the "I will nots" are connected with a "promise to live within the advantages of society" (*GM* II:3). In the third aphorism Nietzsche describes two practices that are "prehistoric" (*GM* II:3) and basic to cooperative sociality—so basic that we see evidence of both in nonhuman animals—which later came to be *interpreted* in terms of the creditor-debtor schema and the idea of equivalence. These practices are punishment and reciprocity. Unfortunately, space precludes me from getting into the details of Nietzsche's argument here; however, it is sufficient for my purposes to show that "I will nots" and debts are not equivalent on his account, and that the former are involuntary obligations that one acquires without making promises.

Nietzsche describes two kinds of creditor-debtor relationships in the second essay, dyadic contractual agreements between two individuals (*GM* II:5), and the communal relationship between the individual qua debtor and society qua creditor (*GM* II:9). On my reading, Nietzsche takes the communal dynamic described in the third aphorism to be fundamental to human sociality. We saw evidence of this in both the third aphorism, where he claims the "I will nots" are "primitive requirements of social co-existence" (*GM* II:3), and in *BGE* 199, where he claims human beings have always lived in social dominance hierarchies.[33] Subsequent to the formation of dyadic contractual agreements, which arose in concert with "the basic forms of purchase, sale, exchange, trade, and commerce" (*GM* II:4), this communal dynamic came to be "interpreted" in terms of two concepts introduced by the creditor-debtor schema, the notion of debt and the principle of equivalence, "the idea that every injury has its *equivalent* in something and can really be paid off" (*GM* II:4). This is what Nietzsche provides analysis of in the ninth aphorism.

"Interpretation" is a term of art for Nietzsche. Interpretations are explanations that "integrate" a practice "into a system of purposes" (*GM* II:12). That is, an interpretation comes into existence by conceptualizing a preexisting practice in terms of specific aims or goals, and in doing so infuses what is otherwise a mere practice—a series of procedures, performed routinely—with meaning. Let's first consider punishment. Punishment consists of a "relatively *permanent*" element, "the practice, the act, the 'drama,' a certain strict sequence of procedures," and a "*fluid*" element, "the meaning, the purpose, the expectation tied to the execution of such procedures" (*GM* II:13). Nietzsche adds that "the procedure itself will be something older, earlier than its use for punishment, that the latter was first placed into, interpreted into the procedure (which had long existed, but was practiced in another sense)" (*GM* II:13). The stable or "permanent" element of punishment, I suggest, is simply that it is a response "to an injury suffered, which is vented on the agent of the injury" (*GM* II:4). This would seem to be the form of punishment we find in the third aphorism, which is described as little more than a natural expression of anger and hostility in response to the violation of an expectation of conformity, and which has the effect of creating and enforcing prescriptive rules, as I argued previously.[34] In the fifth aphorism, on the other hand, punishment is understood to have been "integrated into a system of purposes." It is understood to have become a method for paying off debts, in accordance with the principle of equivalence.

"What is the difference between a mere obligation, a sense that one ought to behave in a certain way, or even that one owes something to someone, and a debt, properly speaking? The answer is simple: money. The difference between a debt and an obligation is that a debt can be precisely quantified."[35] As David Graeber here observes, a debt is an obligation that has a *valuation* attached to it, and this is a point that Nietzsche himself stresses in the fifth aphorism. Creditors administered punishment in a manner that seemed to them "commensurate to the magnitude of the debt," and "exact assessments of value developed from this viewpoint, some going horribly into the smallest details—*legally* established assessments of the individual limbs and areas on the body" (*GM* II:5). This idea of equivalence is completely absent from the third aphorism because it describes an *earlier* practice of punishment, one that makes viable the venture of cooperative sociality but is not yet connected to ideas of fairness.[36] So, the "I will nots" are not conceived by Nietzsche *as debts*—they represent a more primitive form of obligation (i.e., rules).

Now let's consider reciprocity. The fundamental basis of any social venture that aims to secure a common goal, among human or nonhuman animals, is cooperation. However, to cooperate simply means "to act together," and so cooperation covers a wide swath of collaborative activities, ranging from mutualism, a form of "acting together" in which A and B benefit simultaneously (e.g., the coordinated hunting efforts of pack animals), to contractual relations, which are a uniquely human form of cooperation in which A provides some good or service to B on the condition that B promises to reciprocate.[37] Reciprocity is a form of cooperation that lies between mutualism and contracts. To reciprocate means to "give and take mutually" or "to make a return for something."[38] Like the forming of contracts, reciprocity involves a *conditional* exchange of favors, but like mutualism it does not require the making of promises. De Waal, taking as his model Robert Trivers's theory of reciprocal altruism,[39] defines reciprocity as an exchange of favors in which

1. the initial act, while beneficial to the recipient, is costly to the performer;
2. there is a time lag between giving and receiving; and
3. giving is contingent on receiving.

Since giving is contingent upon receiving, reciprocal relations have the same underlying structure as contractual relationships. In fact, de Waal often describes them as quasi-contractual relations governed by implicit promises.[40]

What Nietzsche is describing in the third aphorism is a reciprocal, not a contractual, relationship. First, the reason he claims the "I will nots" are connected to a promise one has made "to live within the advantages of society" is to point out that rule following is a *condition* of receiving its protection. As Nietzsche later observes, the community member "lives protected, shielded, in peace and trust, free from care with regard to certain injuries and hostilities to which the human *outside*, the 'outlaw,' is exposed," and in view of which one has "pledged and obligated oneself to the community" (*GM* II:9). So, in other words, if the community member does not follow the rules, the community will withdraw its protection; that is, he will be liable to punishment. As Maudemarie Clark has noted in reference to this passage, "If you accept the advantages of community life, you are in effect making a bargain with the community, agreeing to go along with the rules that make community life possible."[41]

However, and second, this communal bargain does not imply that society's protection was offered *on the condition* that the individual made a promise to follow the rules. Just like other social animals who reap the benefits of cooperative sociality without making promises, primitive humans enjoyed the benefits of communal life simply by being born into a community, and they received its protection so long as they followed the rules, even if they never made a promise to do so. As we saw previously, the *only* condition that must be satisfied to receive the group's protection is conformity with the "primitive requirements of social co-existence" (*GM* II:3) that make communal life possible. Consequently, the reason Nietzsche says the "I will nots" are "connected" with a "promise one has made" is that one is *obligated* to follow these rules, and so it is *as if* one has made a promise to follow them. The agent, of course, has no option but to follow them, because the community's protection is contingent upon the individual's fidelity to those norms. The "I will nots" are *reciprocal* obligations.

Third, and finally, the promise is *only* implicit in this context because, as I argued previously, what punishment is making the agent "conscious of" is the fact that she is *subject to rules*, or that certain actions are obligatory. In other words, punishment is acquainting her with the basic and indelible reality of obligation as a *social fact*, and it is doing so regardless of whether she ever consented to follow the rules. Consequently, the "I will not" is an *involuntary* obligation. Dyadic contractual relationships, on the other hand, are established by making promises ("Precisely here there are *promises* made" [*GM* II:5]), and so are created by *voluntary* obligations. In this situation a promise *must* be made, unlike in the third aphorism, because promising is a *condition of receipt*: the creditor extends the initial good or service *only if* the debtor produces an expectation of repayment by communicating it. Consequently, it is the making of a promise that gets this whole transaction off the ground in the first place, and so the practice itself assumes that the promisor already has a concept of obligation and a conscience—just as Nietzsche acknowledges in the fifth aphorism.

If this is right, the third aphorism describes a practice of "proto-promising" that must predate the origin of creditor-debtor relationships, a practice we see evidence of in nonhuman animals. The "I will not" is *not* acquired subsequent to making a promise; it is an obligation one incurs merely in virtue of living with others and accepting the "advantages"

of communal life. But since "I will nots" are "requirement[s] of social co-existence" (*GM* II:3), it makes sense to say they are "connected" to a promise. Promising is thus conceived by Nietzsche only as a structural or formal feature of the communal relationship, which is eventually *interpreted* as a creditor-debtor relationship on his analysis, but only subsequent to the advent of the idea of debt and the principle of equivalence. Consequently, the "I will not" is a *reciprocal* and *involuntary* obligation that does not presume the ability to make promises.

Conclusion

This investigation began by raising awareness of a causal dilemma generated by a natural interpretation of the second essay's account of the emergence of conscience and obligation. According to this "contractualist reading," as I called it, Nietzsche takes promising to be a condition of forming contractual relationships, and such relationships are moreover necessary to explain the ability to make promises. I have instead tried to show that he is not committed to this interpretation, since the third and fifth aphorisms articulate two different conceptions of nonmoral obligation. The third aphorism is offered to explain how human beings became "regular" and "predictable" in their behavior by becoming aware of and conforming their behavior to rules, understood to be obligations that one acquires *involuntarily* merely in virtue of living a social form of life. Dyadic contractual agreements, on the other hand, are created on the basis of *voluntary* obligations, by the making of promises, presuming that the debtor already has a conscience and a concept of obligation.

Relative to the scope of the second essay as a whole, my aims here have been quite modest, but I hope they have not been insignificant. I have tried to show that Nietzsche provides a plausible and naturalistic account of the origin of obligation by offering a genealogy of conscience, understood to be the inner voice of a nonhypothetical "ought" inculcated in us through punishment during the morality of custom, prior to the advent of creditor-debtor relationships. Also, I hope to have made sense of the "long history and metamorphosis" (*GM* II:3) the conscience went through prior to it becoming a "memory of the will." If the preceding analysis has been right, this is essentially a history of *involuntary* obligations humans acquired merely in virtue of being the social creatures we ineluctably are.

NOTES

This article began as a dissertation chapter and has benefited from the scrutiny of a number of people at UC Riverside over the last two years. I want to especially thank Zac Bachman, David Beglin, Meredith McFadden, Maudemarie Clark, and Coleen Macnamara for conversations relating to these themes and their feedback on previous drafts. I would also like to thank Pamela Hieronymi, Iain Morrison, and the participants at the 2018 North American Nietzsche Society Conference, particularly Lanier Anderson, Mark Alfano, and John Richardson, for their comments and encouragement. Finally, thanks to Bernard Reginster for writing two excellent and stimulating articles on the conscience; I agree with much more that he has to say in these articles than I disagree with, which I have regrettably been unable to convey here.

1. Citations of Nietzsche's works come from the following translations: *On the Genealogy of Morality*, trans. Maudemarie Clark and Alan J. Swensen (Indianapolis: Hackett, 1998); *Daybreak*, trans. R. J. Hollingdale (Cambridge: Cambridge University Press, 1997); *Beyond Good and Evil*, in *The Basic Writings of Nietzsche*, trans. Walter Kaufmann (New York: Modern Library Edition, 2000); *The Gay Science*, trans. Walter Kaufmann (New York: Vintage, 1974).

2. The memory of the will is an ability to "will on and on something one has once willed" (*GM* II:1). It is exercised by *holding oneself* to normative expectations, whereas the form of conscience that will be my focus here requires only regulating one's behavior in accordance with the expectations of *others*. The memory of the will thus implicates an ability to sustain practical commitment *autonomously*.

3. Nietzsche also refers to this rudimentary form of conscience in the first essay when he remarks that the nobles are "kept so strictly within limits *inter pares*, by mores, worship, custom, gratitude, still more by mutual surveillance, by jealously," and are thus capable of "self-control" and restraint toward one another, although "they are not much better than uncaged beasts of prey toward the outside world which is foreign" (*GM* I:11). "There they enjoy freedom from all social constraint," he says, "in the wilderness they recover the losses incurred through the tension that comes from a long enclosure and fencing-in within the peace of the community; they step *back* into the innocence of the beast-of-prey conscience" (*GM* I:11).

4. See Bernard Reginster, "What Is the Structure of Genealogy of Morality II?," *Inquiry* 61.1 (2017): 1–20; and "The Genealogy of Guilt," in *Nietzsche's* On the Genealogy of Morality: *A Critical Guide*, ed. Simon May (Cambridge: Cambridge University Press, 2011), 56–77.

5. Reginster, "What Is the Structure of Genealogy of Morality II?," 4; and "Genealogy of Guilt," 59.

6. Reginster, "Genealogy of Guilt," 58.

7. Reginster, "What Is the Structure of Genealogy of Morality II?," 4. See also Reginster, "Genealogy of Guilt," 58, 75.

8. Reginster, "What Is the Structure of Genealogy of Morality II?," 4. Aaron Ridley also holds that humans became reliable by making promises. See Ridley, "Nietzsche's Intentions: What the Sovereign Individual Promises," in *Nietzsche on Autonomy and Freedom*, ed. Ken Gemes and Simon May (Oxford: Oxford University Press, 2009), 182. As will become clearer presently, I think Reginster and Ridley here make the mistake of conceiving of reliability of behavior as a condition of *keeping* a promise when it is in fact a condition of being able to *make* one.

9. Reginster, "What Is the Structure of Genealogy of Morality II?," 4.

10. David Owens, "A Simple Theory of Promising," *Philosophical Review* 115.1 (2006): 51–77, 54; Joseph Raz, "Promises and Obligations," in *Law, Morality, and Society: Essays in Honor of H. L. A. Hart*, ed. P. M. S. Hacker and Joseph Raz (Oxford: Oxford University Press, 1977), 210–28, 218; Gary Watson, "Promises, Reasons, and Normative Powers," in *Reasons for Action*, ed. David Sobel and Steven Wall (Cambridge: Cambridge University Press, 2009), 155–78, 156.

11. For more on obligation as a social requirement, see Joel Feinberg, "The Nature and Value of Rights," *Journal of Value Inquiry* 4.4 (1970): 243–6zz0, 244; Phillipa Foot, "Morality as a System of Hypothetical Imperatives," *Philosophical Review* 81.3 (1972): 305–16, 308; and P. F. Strawson, "Social Morality and Individual Ideal," *Philosophy* 36 (1961): 1–17, 5.

12. This is Bernard Williams's definition of obligation, as expressed in *Ethics and the Limits of Philosophy* (Cambridge, MA: Harvard University Press, 1985), 185.

13. Hypothetical imperatives simply recommend the means necessary "to achieving something else that one wants," or what one "ought" to do to be effective in satisfying one's desires. See Immanuel Kant, *Groundwork of the Metaphysics of Morals*, trans. Mary Gregor and Jens Timmerman (Cambridge: Cambridge University Press, 2012), 28.

14. The conscience is invoked to explain how the debtor "impress[es] repayment on his conscience as a duty, as an obligation" (*GM* II:5). Simon May is the only scholar I am aware of who has brought attention to this peculiar feature of the passage. On his reading, Nietzsche simply takes it for granted that the debtor has a "strong" and "ethically charged notion of personal accountability" (*Nietzsche's Ethics and His War on Morality* [Oxford: Oxford University Press, 1999], 56). I have reservations about May's account, since it would seem to assume that at this stage in Nietzsche's story the debtor is already capable of feeling guilt, the "feeling of personal obligation" (*GM* II:8), which creditor-debtor relationships are supposed to explain. That said, I agree with May that the debtor must possess a minimal sense of obligation, as articulated above. I do not take it to be particularly "strong" or "ethically charged," though, for reasons that will become apparent in the third section.

15. I take it this is why Reginster and Ridley take the above passage to be claiming reliability is a *consequence* of promising, rather than a condition of making one. The burden is on me, then, to show that promising is only implicit in this context.

16. John Richardson, *Nietzsche's New Darwinism* (Oxford: Oxford University Press, 2004), 89.

17. Nietzsche does not tell us the origin of the morality of custom, only that it is "prehistoric" (*GM* II:2) and goes back "many millennia" (*D* 14). I agree with Iain Morrisson ("Nietzsche on Guilt: Dependency, Debt, and Imperfection," *European Journal of Philosophy* 26 [2018]: 974–90, 986n7) that prehistoric humans were social and their interactions with one another governed by customs. To the contrary, a number of scholars take Nietzsche to be committed to a presocial state of nature, prior to the formation of states (*GM* II:17). See Reginster, "Genealogy of Guilt," 62; Aaron Ridley, *Nietzsche's Conscience: Six Character Studies from the Genealogy*, (Ithaca, NY: Cornell University Press, 1998), 18–19; Mathias Risse, "The Second Treatise in *On the Genealogy of Morality*: Nietzsche on the Origin of the Bad Conscience," *European Journal of Philosophy* 9.1 (2001): 55–81, 57. Nietzsche takes the earliest tribes to have been "organized according to blood-relationships" (*GM* II:20) and believes customs originated as utility rules, "the experiences of men of earlier times as to what they supposed useful and harmful" (*D* 19). These are what he above refers to as "primitive requirements of social co-existence" (*GM* II:3), which eventually acquired the status of customs, or *Sitte*, a norm that has been passed down through the generations and become a "*traditional* way of behaving" (*D* 9). Accordingly, I will rely on a broad understanding of the morality of custom throughout, taking it to be a primitive form of human social organization that enforced reciprocal obligations and secured reliable interactions among group members by enforcing rules. It is in this capacity that Nietzsche appeals to the morality of custom in the second aphorism.

18. In *Nietzsche on Morality* (London: Routledge, 2002), 229, Brian Leiter similarly maintains that the third aphorism describes "a phenomenon of pre-history: we are discussing what the animal man had to be like before regular civilized intercourse with his fellows ('the advantages of society') would even be possible."

19. A fact of which he was aware, because he attributes this and more complex capacities to animals in *D* 26. There he claims that animals are capable of "objective awareness," of treating themselves as the object of another's gaze, and thus possess the kind of social awareness involved with having a conscience. Like Darwin, Nietzsche held that the human conscience was constructed from various capacities already present in nonhuman animals. See Charles Darwin, *The Descent of Man* (London: Penguin, 2004), 119–51.

20. Frans de Waal, "The Chimpanzee's Sense of Social Regularity and Its Relation to the Human Sense of Justice," *American Behavioral Scientist* 34.3 (1991): 335–49, 337.

21. Frans de Waal, *Good Natured: The Origins of Right and Wrong in Humans and Other Animals* (Cambridge, MA: Harvard University Press, 1996), 96.

22. See de Waal, "Chimpanzee's Sense of Social Regularity," 340; and de Waal, *Good Natured*, 92. See also T. H. Clutton-Brock and G. A. Parker, "Punishment in Animal Societies," *Nature* 373 (1995): 209–16, 211.

23. At least in primates, conformity to a prescriptive rule is not just an instinctual response. De Waal believes their awareness of prescriptive rules is most conspicuous in their efforts to avoid detection for violating norms and when "tattling" on

one another. For some rather comical anecdotes, see de Waal, "Chimpanzee's Sense of Social Regularity," 338–39. See also Christopher Boehm, *Moral Origins* (New York: Basic Books, 2012), 106, 116–29.

24. H. L. A. Hart, *The Concept of Law* (Oxford: Clarendon, 1961), 86.

25. Reginster, "What Is the Structure of Genealogy of Morality II?," 4.

26. Williams, *Ethics and the Limits of Philosophy*, 185.

27. Williams, *Ethics and the Limits of Philosophy*, 187.

28. Paul Katsafanas (*The Nietzschean Self* [Oxford: Oxford University Press, 2016], 89–90) observes that scientists and philosophers in the nineteenth century typically distinguished instinctual actions or responses from learned behaviors, taking the former to be unreflective and the latter to require awareness of an end or goal (e.g., avoiding punishment). As we can see, it is plausible to think Nietzsche had the same distinction in mind in the third aphorism of *GM* II.

29. Kant, *Groundwork*, 44.

30. See Maudemarie Clark and Brian Leiter, "Introduction," in *Daybreak*, trans. R. J. Hollingdale, (Cambridge: Cambridge University Press, 1997), xxx.

31. Foot, "Morality as a System of Hypothetical Imperatives," 309.

32. The herd instinct is a disposition to conform one's behavior to that of others (*GS* 116). The "herd instinct of obedience" is a disposition to give precedence to the commands of those who have rank within dominance hierarchies. Stanley Milgram's famous and disturbing "obedience study" illustrates just how ingrained this disposition still is in human beings today. See Milgram, "Behavioral Study of Obedience," *Journal of Abnormal and Social Psychology* 67 (1963): 371–78.

33. We also see evidence in *GM* II:8, which on the face of it seems to present evidence to the contrary. Specifically, Nietzsche remarks that the economic instruments associated with the creditor-debtor schema, along with the "rudimentary" ideas of "exchange, contract, guilt, right, obligation, compensation" were "*transferred* [. . .] onto the coarsest and earliest communal complexes" (*GM* II:8). This presumes that humans were living in groups prior to the advent of creditor-debtor relations, despite Nietzsche's claim that they are "older even than the beginnings of societal associations and organizational forms" (*GM* II:8). Space precludes defense here, but I believe he takes these "associations" and "forms" to be *government* or *political* organizations.

34. As noted by Maudemarie Clark ("Nietzsche on Free Will, Causality, and Responsibility," in *Nietzsche on Ethics and Politics* [Oxford: Oxford University Press, 2015], 75–96, 93), Nietzsche appeals to a similar notion of communal "punishment" in the ninth aphorism, prior to the idea of equivalence. He also acknowledges that "punishment" is not uniquely human: "Seeing-suffer feels good, making-suffer even more so—that is a hard proposition, but a central one, an old powerful human-all-too-human proposition, to which, by the way, even the apes might subscribe: for it is said that in thinking up bizarre cruelties they already abundantly herald and, as it were, 'prelude' man" (*GM* II:6).

35. David Graeber, *Debt: The First 5,000 Years* (Brooklyn, NY: Melville House, 2011), 21.

36. Equivalence is "the oldest and most naïve canon of moral justice" (*GM* II:8).

37. Stephen I. Rothstein and Raymond Pierotti, "Distinctions among Reciprocal Altruism, Kin Selection, and Cooperation and a Model for the Initial Evolution of Beneficent Behavior," *Ethology and Sociobiology* 9.2–4 (1988): 189–209, 198.

38. Rothstein and Pierotti, "Distinctions among Reciprocal Altruism," 198.

39. See Robert Trivers, "The Evolution of Reciprocal Altruism," *Quarterly Review of Biology* 46.1 (1972): 35–57.

40. See Frans de Waal and Jessica Flack, "Any Animal Whatever: Darwinian Building Blocks of Morality in Monkeys and Apes," *Journal of Consciousness Studies* 7.1 (2000): 1–29, 19–20.

41. Maudemarie Clark, "Nietzsche's Immoralism and the Concept of Morality," in *Nietzsche on Ethics and Politics*, 23-40, 36.