NIETZSCHE’S ANSWER TO THE NATURALISTIC FALLACY:

LIFE AS CONDITION, NOT CRITERION, OF MORALITY

Donovan Miyasaki

Wright State University

Many scholars interpret Nietzsche’s moral philosophy as a form of ethical naturalism, usually a variation of perfectionism or virtue ethics in which excellence correlates with the natural value of power. On the ethical naturalist reading, Nietzsche believes his evaluative judgments of noble and base, higher and lower, healthy and decadent are truer or more reasonable than those he critiques, a truth that ultimately rests on objective claims about the natural world.

Nietzsche’s account of organic life as will to power certainly tempts toward such a view, since it seems to imply that life has an intrinsic aim according to which its success can be measured. Life, he says, is in its very nature a will toward power, toward quantitative growth through expansion of power over one’s environment. Organic life is “*essentially* appropriating, injuring, overpowering, oppressing, being harsh, imposing your own form, incorporating, and exploiting (or depleting)” (*BGE* 259).

But this is misleading. First, it implies that life and growth of power are identical, which they cannot be if, as *Zarathustra* tells us, life “sacrifices itself for power” (*Z* II: 12). Second, it suggests that power is a teleological purpose. But there’s no particular state of power, such as survival, reproduction, or domination, at which life aims. Rather, life tends toward a mere “discharge of strength” (*auslassen*)—not an accumulation but a *venting*, a release that only accidentally produces outcomes like growth as its “most common and indirect consequences (or *Folgen*, *BGE* 13)*.*”Power doesn’t *lead* life but *follows* it, like a skin that it casts off or like Zarathustra’s image of the “falling and perishing leaves” (*Z* II: 12).

Consequently, organic life cannot, as ethical naturalist readings claim, provide a criterion for evaluating life. An organism’s value doesn’t correspond to its power or growth, for they aren’t aims that it essentially seeks or ought to achieve.

However, *conscious animal* life does produce an evaluative criterion. For animals possess affectivity, making them aware of their tendency toward power. In animal *pleasure and pain*, life evaluates its success. And so animal life contains a *universal* criterion for evaluating life: the *happiness* that coincides with power’s growth.

This, in turn, provides Nietzsche with criteria of *health* and *decadence*: when “life is *ascending*, happiness is identical to instinct” (*TI*, Problem 11). *Animal health* is, then, the coincidence of happiness and growth in power, a state in which an animal’s instincts direct it toward conditions conducive to growth, while *decadent* instincts direct away from such conditions.

But this universal value of animal health as the coincidence of happiness and growth is not, as ethical naturalism requires, a criterion of *moral* evaluation. If we equate animal health with moral value, we commit the naturalistic fallacy of defining moral properties according to factual ones. Maybe it’s in our nature to find happiness in power, but goodness isn’t conceptually reducible to power. We can still, following G. E. Moore’s “open question argument,” meaningfully ask: is power really *always* good? In the broader Humean sense of the fallacy, we can still wonder how the normative conclusion follows: why *ought* we maximize our happiness? Why condemn someone who’s content with middling happiness? Are Zarathustra’s last men, for example, those all-too-contented cows, really morally *obliged* to be otherwise? Are the *Genealogy’s* tender little lambs any more *blameworthy* than its birds of prey?

Even more troubling for the ethical naturalist is the fact that the value of animal health is deeply at odds with morality: as *appropriative*, healthy animal life increases power by exploiting others, tending toward *instrumental, dominating* relations, and as *accumulative*, life is insatiable in its appropriation, tending toward *conflicting, violent* relations. Far from providing a criterion *for* morality, animal health is the principal *obstacle* *to* morality, a seemingly insurmountable, natural basis of all moral conflict.

At this point interpretations diverge. Ethical naturalists endorse a more moderate conception of will to power capable of moral refinement, while the anti-realists (such as Leiter, Nehamas, and Hussain) suggest Nietzsche’s values are not meant to be taken as true, abandoning any attempt to ground them.

There’s a degree of truth in both views, but both neglect the deep differences between Nietzsche’s conceptions of animal and human health. In the *Genealogy*, he depicts humanity as a sick animal transformed by traumatic environmental change. Instincts once in tune with animal conditions of growth suddenly become useless in the new environment of civilization. The animal form of health, the harmony of happiness and power, no longer applies.

However a new, distinctly human kind of health develops through what Nietzsche calls the “social straightjacket” of “conventional morality” (*Sittlichkeit der Sitte*): the sick human animal is transformed into a *conventional moral agent* through social training and disciplinary systems of punishment (*GM* II: 2). The lost animal harmony of instinct and happiness is restored in a second nature, whose socially-shaped drives have been harmonized with humanity’s new environment. Health no longer coincides with domination of one’s environment, but instead *conformity* *to* it—paradoxically, with a *decline* in growth and power.

However, the invention of moral agency raises a problem: how is this subordination to convention possible if the very essence of life is insatiable, violent domination of one’s environment? Nietzsche’s answer is that moral agency doesn’t directly oppose will to power, but redirects it toward internal objects of exploitation: one’s own animal instincts.

And so moral agency is a new kind of health analogous to the animal form. Like animal health, it involves a growth in power, but over the internal rather than external world. And like animal health, it achieves a unity of happiness and instinct: for humans, social cooperation is a new environmental condition of power, and moral agency trains individuals to find happiness in forms of power compatible with those new conditions. It’s the same concept of health, but adapted to humanity’s new environment.

However, Nietzsche underemphasizes this profound conceptual shift: if the will to power can be satisfied in the qualitative moral feeling of self-command rather than the quantitative achievement of domination, then life was *never* really about growth at all! Quantitative growth is just another accidental consequence of will to power, not its essential aim. On the human level, then, the growth of power is no longer a universal criterion of value. It doesn’t measure human happiness, nor does it, as the ethical naturalist requires, provide a *normative* criterion of *moral* *value*.

This, in turn, forces us to reconsider Nietzsche’s conception of life and reconceive human health in a way compatible with the internal, qualitative form of will to power found in moral agency. What then are the distinctively human forms of health and happiness? In *Twilight*, he tells us that the good is “everything that heightens [*erhöht*] people’s *feeling of power*, the *will to* power, power *itself*” (*A* 2). In this series of causally-related properties, each item is a means to another, with *feeling*, not power, as the final end: in order to *feel* that one has gained power, one must *possess* some power, and in order to *possess* power, one must have the *will* to gain it. Human life as will to power is not fundamentally a tendency toward power but toward the *feeling* of power.

Nietzsche has always emphasized this. In *The Gay Science*, he speaks of his “doctrine of *power-feeling*” (*GS* I: 13) and in *Beyond Good and Evil* he emphasizes that we aim not at growth or increased power but “more precisely”—his words—“the *feeling* of growth, the *feeling* of increased power” (*BGE* 230).

Only recently have more scholars begun to emphasize this shift toward feeling in the human form of the will to power. For example, Reginster and Katsafanas have stressed that the will to power is as much about the *activity* of overcoming obstacles as it is about the *goal*. But this still overstates the importance of overcoming, which is just another accidental consequence, not an aim. As a *feeling of* power, the will to power aims not at the activity of *overcoming* resistances, but instead at the *activity* *of resisting* for its own sake.

This distinctly human form of will to power in turn provides us with a uniquely human form of health: the coincidence of happiness with conditions—not of quantitative growth—but of *qualitatively heightened feelings* of power. This qualitative shift explains why Nietzsche’s preferred term for human enhancement comes from the word “heightening” or “elevation.” Healthy individuals seek power not as quantitative increase, but as the heightened qualitative intensity of resisting activity.

And that feeling of power is optimized in relations of equality. The first defining feature of health is the search for resistance: “every growth reveals itself in the seeking out of a powerful opponent”(*EH* “Wise” 7). However, the second and more crucial defining feature is specifically preferring *equal* resistances, ones likely to thwart overcoming: “The task isn’t at all about mastering resistances in general but instead those that demand your whole strength, suppleness, and skill with weapons, – opponents that are your equals”(*EH* Wise 7).

And so human health tends toward establishing and preserving precarious relations of equal power and resistance. Nietzsche has identified a universal criterion for evaluating human life that is, in contrast to animal health, not essentially in conflict with morality, one that we can pursue without seeking to dominate others. Indeed, we maximize the feeling of power in resisting activity precisely by preventing domination, by sustaining equal contests rather than definitively or finally winning them.

In this respect, then, the ethical naturalist is right: taking human nature as the ground of moral evaluation doesn’t, after all, require justifying domination, since they’re merely accidents of the unique environmental conditions of pre-human life, and not the necessary aim of all living things.

But this still doesn’t solve the problem of the naturalistic fallacy. Pursuing human health in the form of equal resistance may not necessitate moral conflicts, and it may be universally desirable as a source of happiness, but Nietzsche still can’t conclude that we *ought* to pursue health more successfully than we already happen to. It’s still not a criterion of *moral* value with obligating authority over those who fail to realize it.

In this respect, the anti-realist is right. Life and health are not *criteria* of morality. However, what the anti-realist misses is that Nietzsche’s views about health are still deeply connected to morality. Health isn’t a *criterion* but rather a *condition* of morality: it’s necessary for the ability to effectively practice any kind of morality at all, even though it doesn’t require any particular set of moral principles or values.

Now, as we’ve seen, this distinctly human form of health originates with the social production of conventional moral agency. Because conventional moral agency is subordinated to social mores, it’s tempting to assume it’s an example of *decadence*, and that only its successor, the so-called “sovereign individual” demonstrates true health.

However, we can now see that’s not the case. Remember that for Nietzsche, health is a unity of instinct and happiness in conditions that heighten power. Human health develops out of moral agency rather than other forms of agency, because that’s the form of selfhood that maximizes the feeling of power in a way most compatible with the distinctly social human conditions of power.

It does so for two reasons. First, moral agency, like any other form of agency, unites an individual’s various drives under a single stronger drive in the service of a single value or aim. In doing so, it produces unity of will, an “enduring, unbreakable will” that is the foundation of the “prerogative to make promises” and the capacity to act in the face of resistance—thus, to experience a greater feeling of power in relation to external resistances (*GM* II: 2).

Second, unlike *non-moral* forms of agency, *moral* agency also optimizes the conditions for an individual to feel power in relation to internal resistances. For moral agency unites our consciously affirmed values with our strongest drives, so that we experience successful action not merely as obedience to our drives, but as a feeling of freedom in self-command. Our apparent freedom of the will is purely an affect, a “state of pleasure of one who commands and, at the same time, identifies himself with the accomplished act of willing” (*BGE* 19).

Contrast, for example, the non-moral agency of an addict. Her instincts are also united by a single organizing drive, but she experiences that unity as slavery rather than freedom. Or contrast the simple agency of an animal with no moral conflicts: without strongly opposing drives to overcome, it has no internal sources of resistance and the feeling of power.

So, it is the *form* of moral agency, not the *content* of its values, that optimizes human health. It does by maximizing the feeling of power against both *internal* and *external* resistances, against internal resistance by combining opposing drives, and against external resistance by organizing them hierarchically toward successful action.

From this we can draw two surprising conclusions. The first is that there’s nothing objectionable in principle about conventionality. If health consists in the hierarchical organization of the self under commanding drives that are aligned with consciously affirmed values, then the content of those values is irrelevant, as long as they still organize the drives toward effective action in accordance with those values.

Consequently, Nietzsche’s ideal of health is, just as he claims, a truly “immoralist” one. He avoids the naturalist fallacy by refusing to draw moral conclusions, as well as by refusing to give any specific normative content to his conception of moral agency. Against the ethical naturalist reading, we’re not *obligated* to pursue or maximize health, nor are we obligated to pursue our happiness in any particular form of power, such as physical strength, intellect, or artistic creation. As he says about “giving style to one’s character”: “whether the taste was good or bad means less than one may think; it's enough that it was one taste!” (GS 290). From the standpoint of life as a will to power, any source of the feeling of power will do.

However, against the anti-realist readings, Nietzsche’s values are not subjective matters of taste. Every human being has a rational self-interest in adopting his ideal of health, since it optimizes our tendency toward the feeling of power, maximizing our happiness.

Our second surprising conclusion is that health is a basic precondition of all moral agency. One must first be healthy in order to successfully be any kind of moral agent at all. Consequently, there’s no such thing as *decadent morality*, for there are no decadent moral agents to enact it. All effective morality is by definition healthy. And it’s this correlation of decadence and the lack of effective moral agency that grounds Nietzsche’s solution to the naturalistic fallacy.

Against the anti-realist reading, Nietzsche doesn’t just sidestep the naturalist fallacy. To do so would make his ideal of health morally trivial, unable to resolve moral conflicts. But rather than endorsing moral norms in order to prevent such conflicts, Nietzsche’s ideal instead resolves them at their source.

As we’ve seen, the source of moral conflict is organic life’s tendency toward domination. But as we’ve also seen, the human form of health optimized in moral agency redirects life toward the feeling of power in the activity of resistance for its own sake. Consequently, the greater the degree of moral agency, the more the individual’s happiness coincides with conditions of sustained, equal resistance, with contest rather than conquest. Happiness is maximized in the achievement of moral agency, which in turn reduces the source of moral conflict: the desire to increase power through domination.

Admittedly, since the content of a moral agent’s values is irrelevant, it’s possible in principle for moral agents to explicitly organize their drives under aims that affirm moral conflict—for example, those who adopt racist or nationalist ideologies. However, in practice, these forms of moral agency are ultimately self-defeating, because strong moral agents will find greater happiness in self-discipline and relations of equality than in fully realizing their professed aims. Like athletes whose devotion to winning overtakes their pleasure in the game, their success is self-undermining: if they are to preserve their motivation, they must find worthy contests and equal competitors.

In other words, moral agents who affirm violence or domination, do so from a *lack* of strong moral agency, rather than *from* it. The greater the disunity in their drives, the more ineffective they are in their attempt to act against external resistances, causing them to experience the power and agency of others as the source of their own feelings of impotence. As in the *Genealogy*’s account of slave morality, it is precisely *weak* moral agency that creates an incentive to overpower, weaken, and harm others. Moral conflict is based primarily in failed moral agency, rather than in the content of moral agency—in an inability to integrate values, rather than possession of the “wrong” values.

And while strong moral agency *reduces* the incentive toward moral conflict regardless of its content, weak moral agency *increases* it regardless of content. Even when their values explicitly condemn domination, weak moral agents are motivated to exercise those values in violent and dominating ways, just as, in the *Genealogy*, the priests use of the psychology of guilt to harm others, despite their explicit commitments to love, compassion, and mercy.

And so Nietzsche’s ideal of human health, although not a moral ideal, is far from morally trivial. If human life is essentially a tendency toward the feeling of power in equal resistance, then strong moral agency, regardless of its content, tends toward the reduction of moral conflict. And if that is so, then moral philosophy is deeply mistaken in its attempt to reduce moral conflict primarily through moral content, rather than through the general promotion of moral agency of every kind.

In this way, Nietzsche’s immoralism not answers the naturalist fallacy, but suggests an overlooked but novel criticism of our entire moral tradition: that its greatest failure is its faith in the efficacy of moral content to solve moral conflict, its endless preoccupation with determining which values are the correct ones.