

Received: 1 October 2021 Accepted: 5 November 2021

DOI: <https://doi.org/10.33182/agon.v15i3.1913>



Review of John Richardson, Nietzsche's Values

Paul Katsafanas¹

Abstract

This article examines John Richardson's Nietzsche's Values. Richardson's book is systematic in the very best sense. He patiently works through the apparently contrary claims that Nietzsche makes about each topic pertaining to values. In each chapter, Richardson shows that these apparently contrary claims are not only reconcilable, but are interlocking: they support one another, constituting an impressively unified analysis of the human condition. By the end of the book, Richardson produces a comprehensive analysis of Nietzsche's thought on values, will to power, life, consciousness, agency, freedom, culture, and religion. While the book is impressive, I critique Richardson's treatment of four points. Section One argues that the form of internalism that Richardson attributes to Nietzsche is somewhat underspecified. Section Two asks whether Richardson's version of internalism can account for the immense distance between what we do value and what we should value. There, I also raise some questions Richardson's interpretation of will to power. Section Three suggests that Richardson's reading of Nietzsche's ethics is much closer to constitutivism than he acknowledges, and that fully endorsing constitutivism would resolve some of the problems that Richardson's account otherwise faces. Section Four argues that Richardson's distinction between animal drives and socially induced drives is problematic.

Introduction

It's all too common for Nietzsche scholars to emphasize one side of Nietzsche's thinking on a topic while excluding or minimizing discussion of another side. Nietzsche's texts are so rich, so full of disparate and even apparently conflicting claims, that it's easy to be drawn so deeply into one vein of text that you lose sight of others. What's remarkable about Richardson's book is that it never falls into this trap. Richardson's work is systematic in the very best sense. He patiently works through the apparently contrary claims that Nietzsche makes about each topic. In each chapter, Richardson shows that these apparently contrary claims are not only reconcilable, but are interlocking: they support one another, constituting an impressively unified analysis of the human condition. This requires meticulous and sometimes painstaking analysis: the book is *very* long, and at times the reader might wish that Richardson could reconcile the topics more quickly. But the

¹ Paul Katsafanas, Boston University, United States. E-mail: pkatsa@bu.edu.

Paul Katsafanas is Professor of Philosophy at Boston University. He is the author of *Agency and the Foundations of Ethics: Nietzschean Constitutivism* (Oxford, 2013), *The Nietzschean Self: Moral Psychology, Agency, and the Unconscious* (Oxford, 2016), and *Devotion: Nihilism, Fanaticism, and the Longing for Invulnerable Ideals* (Oxford, forthcoming). He has published approximately 30 papers on Nietzsche, moral psychology, and ethics.



writing is refreshingly clear and the interpretations are illuminating. By the end of the book, Richardson produces a comprehensive analysis of Nietzsche's thought on values, will to power, life, consciousness, agency, freedom, culture, and religion. No small feat!

I will try to summarize Richardson's interpretation of Nietzsche. We can start with a point about consciousness. Nietzsche is fascinated by non-conscious processes and thinks that we tend to overestimate consciousness' scope and import. Nonetheless, Nietzsche sees consciousness as transforming us in profound ways. Although many of the ways in which we consciously conceive of ourselves are mistaken, they nonetheless make a difference in what we are. These errors, which resound through our self-conceptions and become incorporated into our pre-reflective ways of relating to the world, are in some cases disastrous; they need to be studied, rooted out, and corrected. And one facet of this is especially important: we need to examine our *values*.

Nietzsche's chief philosophical preoccupation is values. We can speak of three types of valuing. First, there are "body-values," which Richardson identifies with the ends induced by our drives. Second, there are "human values," which are the criteria that reflective agents appeal to in deciding what to do; they are based on social rules and customs and designed for the good of the herd. These human values are linguistically formulated and are reflexive in a way that bodily values are not. Finally, there is a potential for creating "superhuman" values, which would require self-understanding and would be more oriented toward the unique goods of the individual who creates them.

Nietzsche rejects the idea that values are part of the fabric of the universe, instead maintaining that all values arise from acts of valuing. Nonetheless, he thinks that certain values are criticizable whereas others are not. His standard for critiquing values is internalist: any claim about what we *should* value must be based on a claim about what we *do* value. Nietzsche wants to destabilize many of our values, in part by showing that when we understand how they arise and function, they are disrupted. But not all values would be destabilized in this way. Nietzsche thinks that some values could be maintained together with an understanding of their nature, source, and effects. These are the values that he champions; they would be superhuman values.

But we don't yet have superhuman values; we only have traditional, human values. These traditional values have led us to nihilism, which takes two basic forms. One form of nihilism is the collapse of values (leaving us without any values). Another form of nihilism is having life-negating or anti-life values. Nietzsche wants to overcome both forms of nihilism. He charts a path to better values, values that we can hold in light of an understanding of what they are, what they do, and where they come from. In striving for these superhuman values, we would attain a deeper form of freedom.

Part of attaining this deeper form of freedom is becoming more reflective, enabling us to root out external influences, both at the level of bodily drives and at the level of social customs and values. But when we are studying these values and gaining a deeper understanding of them, what do we look for? How does acquiring information about the nature, source, and effects of values help us to assess them? It does so by relating them to *life*. Richardson argues that life is Nietzsche's key evaluative concept: we should reject values that conflict with or undermine life, instead embracing values that are life-affirming. Life-affirming values bear a special relation to will to power; power is the particular form



of life that Nietzsche endorses. So Nietzsche's account of life and its normative status depends on an account of will to power. Will to power is analyzed in terms of growth: to will power is to incorporate competing forces into one's own activities.

Although attaining freedom and crafting life-affirming values have individualistic or egoistic dimensions, they also have a social component: the free individual, in crafting values that are in some sense her own, would also serve (in the exemplary case) to incorporate these values into the social realm. After all, Nietzsche doesn't just want isolated individuals to create new values for themselves. He also wants life-affirming values to spread through culture, transforming our social and cultural life. Nietzsche speculates that in order to generate life-affirming values that function as culturally effective phenomena, we need to create new religions. Yet these "religions" are not religions in the ordinary sense: they do not make claims about the existence of divine entities. They instead cultivate certain moods or orientations toward life, which are manifest in social and individual practices. While traditional religions have generated and sustained deleterious moods and orientations, Nietzsche envisions an affirmative religion that would produce a form of joy in which we say Yes to everything, feeling and seeing the world as divine. This would result in a Dionysian affirmation of life.

That, in essence, is the picture that emerges from Richardson's book. I find it very hard to critique Richardson's book because I agree with almost everything he says. Those are *exactly* the topics that concern Nietzsche and I accept a version of each of Richardson's claims about them. Of course, I have disagreements with some of the details, and I would phrase certain points differently, but in broad outlines I think Richardson's reading of Nietzsche is exactly right. But, in the spirit of pushing against some of these details, I will raise four points. Section One argues that the form of internalism that Richardson attributes to Nietzsche is somewhat underspecified. Section Two asks whether Richardson's version of internalism can account for the immense distance between what we do value and what we should value. There, I also raise some questions Richardson's interpretation of will to power. Section Three suggests that Richardson's reading of Nietzsche's ethics is much closer to constitutivism than he acknowledges, and that fully endorsing constitutivism would resolve some of the problems that Richardson's account otherwise faces. Section Four argues that Richardson's distinction between animal drives and socially induced drives is problematic.

Nietzsche's metaethical views

Nietzsche makes two claims that might seem incompatible: he wants to maintain both that

- i. all values are created by valuing creatures and that
- ii. we ought to reject certain values while embracing others.

Claim (i) might suggest a form of relativism, according to which we have no grounds for assessing disparate sets of values. Claim (ii) might suggest a form of absolutism, according to which we have some objective standard in terms of which we can assess disparate sets of values. Let's see how Richardson navigates this potential dilemma.

Richardson begins by distinguishing "full value realism" and "full value perspectivism". Full value realism holds that "what's good for person P is (completely) independent of what P does value" (9), whereas full value perspectivism holds that "what's good for

person P is (precisely) what(ever) P does value" (10). Richardson claims that Nietzsche denies both of these views. Nietzsche's critiques of value realism are obvious; he repeatedly denies that values would exist independently of creatures who value (see, for example, D 120, GS 301, Z I.15, and TI IX.37). Nietzsche's rejection of full value perspectivism, though perhaps less often appreciated, is also clear. For

too directly tying P's good to P's valuing would rule out any effort to persuade the person that what's good for her or him is something *other* than what he or she now values. That is, it would rule out any effort to persuade by *reasons*, rather than by coercion or propaganda. But Nietzsche very much wants to give reasons in favor of his values—reasons why we readers should change what we value by adopting his values. (10)

This seems to me correct. Nietzsche wants to reject full value perspectivism (or, in more familiar terms, strong value relativism) in favor of something more robust.

Nietzsche must accept some intermediary position, avoiding both of these extremes. Richardson describes Nietzsche as follows: he "privileges his own values by their satisfaction of epistemic constraints" (34). In particular,

Nietzsche justifies his values by direct appeal to the values we already have. He tries to point out values we have without noticing them. The "ought" is supplied not from outside but by what the person values already. (34)

In more familiar language, Nietzsche is some kind of *internalist*: he thinks that what we *should* value depends, in some way, on facts about what we *do* value (10-11).

I agree with Richardson that Nietzsche endorses some form of internalism. But I think we can raise questions about the particular form of internalism that Richardson attributes to Nietzsche. Typically, internalism is specified in terms of a connection between *motives* and *reasons*. For example, Bernard Williams' famous version of internalism claims that we have reason to A iff there exists a sound deliberative route from our motives to A-ing (for example, I have reason to drink this cup of coffee iff I have a motive that would be fulfilled by drinking it). Richardson's version is different: rather than basing claims about *reasons* on claims about *motives*, he bases claims about what we *should value* on claims about what we *do value*. I have two worries about this.

My first worry is that the value-based version of internalism is going to be *far* more complicated than the standard versions of internalism. There is a difference between having a reason to A and valuing A. I can have a reason to drink the stuff in front of me without valuing the stuff in front of me. I can have a reason to do things that I don't value, such as throwing out the garbage. And I can value something without having any immediate reasons pertaining to it. For example, I value the beauty of Arches National Park but don't see any way in which this generates any reasons for me right now. So straightaway we can see that complications are introduced by focusing on values rather than more primitive, less complex psychological states.

This might make us skeptical of Richardson's value internalism. Why think that claims about what we should value have to be grounded in claims about what we *do value*? Why



wouldn't it be enough to instead ground claims about what we should value in claims about our attitudes, for example?

That brings me to my second worry. Can Richardson's value-based version of internalism really account for Nietzschean claims about what we should value? Consider an example. Let's imagine a committed religious ascetic who thinks that sex is disvaluable. Let's further suppose that he has strong aversions to sex: when contemplating sex, he feels guilty, defiled, corrupted; when in sexually charged situations, he tries to escape; and so on. I don't think there's any plausible sense in which this individual can be said to value sex. Nonetheless, we can imagine that in counterfactual situations some of his attitudes would support a valuation of sex. For example: perhaps if he shed his dogmatic religious beliefs he would gradually come to see sex in a new way. He would be less conflicted; his desires wouldn't be experienced as sinful and hence wouldn't give rise to negative self-regarding affects. So, changing some of his attitudes (some beliefs, in this case) would result in his being less conflicted. We might describe this by saying that a more fully informed version of the agent would value sex. And this might lead us to say that although the agent doesn't value sex, he should. But this claim about what he should value isn't based on what he does value. It's based on a more complex set of claims about his attitudes (including his beliefs, desires, and affects) and their interrelations.

I'm not sure that Richardson would disagree with any of this. But it does make me think that starting with claims about what we *value* is less perspicuous than starting with more specific claims about our attitudes and their interrelations. I also don't see why we should read Nietzsche as grounding claims about what we should value exclusively on claims about what we do value, rather than on a broader set of psychological claims about our drives, affects, beliefs, and motivational states.

The gap between what we do value and what we should value

Let's now consider a different potential problem with Richardson's interpretation of Nietzsche on value. Nietzschean internalism is supposed to entail that we have reason to change our values. As Richardson puts it, Nietzsche

needs to persuade his readers, and even himself, that his values 'should' be valued. When he values, for example, power, he can't mean *just* that he values it... He needs a way to prioritize [power and other values] over other valueds. He needs to give us, his readers, *reasons* to adopt these values himself. Yet these reasons can't claim that his values are real or external goods. His reasons must appeal to the values we already have; they must show how our own values give us these reasons to change our values. (33)

If the does-value/should-value connection were very close—if what I should value involved only minor departures from what I do value—then a Nietzschean version of value-internalism would be easy enough to understand. But Nietzsche thinks that what I should value can be *very* far from what I do value. So Richardson needs his value-internalism to reconcile the following two claims:

- i. What I should value is based on what I do value.
- ii. What I should value is sometimes *very far* from what I do value.

The problem is that these push in contrary directions. (iii) makes it seem as if there can be only local corrections in our values, whereas claim (iv) denies this. Claim (iv) makes it seem as if what we should value will bear only the most tenuous connection to what we do value, whereas claim (iii) suggests the opposite.

It's clear that Nietzsche thinks there can be massive divergences between what a person does value and what they should value; this is a constant refrain in his works. To cite just two examples: "it is my contention that all the supreme values of mankind [...] are symptomatic of decline, nihilistic values" (A 6). "Value judgments have been stood on their heads and the concepts of 'true' and 'false' are of necessity reversed: whatever is most harmful to life is called 'true'; whatever elevates, enhances, affirms, justifies it, and makes it triumphant is called 'false'" (A 9). So Nietzsche thinks most of our current values are mistaken: what we should value is very different than what we do value. Nonetheless, if Richardson is right about Nietzsche being an internalist, what we should value must somehow be based in what we do value. How can this be?

To see how this might work, we first need to figure out what Nietzsche thinks we should value. Richardson notes that Nietzsche has a clear answer to that question: Nietzsche thinks we should value life or power. The passages I've just quoted—especially A 9—make this clear, but they are far from alone. As Richardson extensively documents in Chapter 2, Nietzsche consistently appeals to either life or power when evaluating values. Life "lies at the very center of his thinking" and constitutes "the clear fulcrum of most of his defenses of his values" (39). Life "is Nietzsche's *principal* justification for his values: life gives the main criterion by which he carries out his 'revaluation of values'" (39). This point seems to me exactly right and, as Richardson notes, it is astonishing that it is so often ignored; when you look at Nietzsche's texts, especially from *Zarathustra* onwards, it takes a certain kind of willful ignorance to read Nietzsche as doing anything other than appealing to life as his principal evaluative standard.

Of course, the appeals to life are bound up with appeals to power: "life is ... will to power" (KSA 7:7[9]); "life itself is *will to power*" (BGE 13); "the will to power is simply the will of life" (GS 349); we have mistaken "the essence of life, its *will to power*" (GM II.12). So one thing that we need to sort out is the connection between life and power. Richardson offers a helpful discussion of whether one of these—life or power—is supposed to be primary. Nietzsche might explain the value of life in terms of the value of power (roughly, life is valuable because it manifests power); or he might explain the value of power in terms of the value of life (roughly, power is valuable because it is an essential feature of life). Which is correct? Richardson thinks that Nietzsche's answer is complicated. Life is more basic than power; the value of power is explained in terms of the value of life. But there are different "levels or degrees of life," and power picks out the particular level or degree of life that Nietzsche prizes (56). So what Nietzsche values most highly is not really life as such but a particular configuration or form of life, which he describes in terms of will to power (56-7). This seems to me to be an especially helpful way of framing the issue.

We still need to know what Nietzsche means by power. Richardson claims that "Nietzsche's core idea of power is *'growth in control'*" (57). But this needs explanation as well. Not every form of growth counts as growth in control. Rather, Nietzsche



has a quite particular kind of growth in mind, one that involves—but is also more than—bringing into control something alien and competing... For the growth that Nietzsche has in mind, the competing force is not simply destroyed...but incorporated by having its different abilities adapted to serve the living thing's projects. It's by this incorporation [*Einverleibung*] of something foreign that the old life is disrupted and outgrown, and the jump to a new level occurs. (57)

To clarify this, we can distinguish it from Reginster's influential claim that will to power is the will to the activity of overcoming resistance (Reginster 2006: 127). Richardson thinks that overcoming resistance is part of will to power. But you can overcome resistance without growing, without incorporating the thing that resists. You could, for example, engage in overcoming a continuous series of very minor resistances. While this would fulfill Reginster's criterion, it wouldn't constitute growth. Thus, Richardson's notion of will to power is more demanding than Reginster's. Richardson puts his own view this way: "Incorporation is a matter of coming-to-control another and contrary will by inducing it to serve one's activity. By building it into that activity, one grows *in* that activity" (58).

I think there is much that's illuminating here; the whole chapter, with its extensive discussions of the passages in which Nietzsche associates power with growth, is invaluable. But there's a potential problem: the more demanding we make the notion of will to power, the less plausible it will be that this is an essential feature of life. I will highlight one aspect of this. Richardson's notion of will to power focuses on incorporating opposing *wills*. But why do the challenges that we grow by surmounting have to take the form of wills? Suppose I set myself the challenge of writing a novel, or climbing a mountain, or painting a portrait. I spend years struggling at the task, gradually growing more adept and successful. Eventually, I produce something (a novel completed, a peak reached, a portrait painted) that satisfies me, something that I couldn't have done at the outset. This looks like a paradigmatic case of growth. And yet it's not clear how these tasks would involve incorporating opposing wills. They seem to involve growth in skill or ability rather than growth in the number of wills incorporated.

To be clear, I don't doubt that incorporating (or sublimating) opposing wills is *one form* of growth in power. It certainly is. Nor do I doubt that Nietzsche values this form of growth. He clearly does, mentioning it in relation to Goethe and others. What I doubt is that this is the *only* way of expressing power. Nonetheless, I do think that there is something right about Richardson's connection between power and growth. Nietzsche draws a very tight connection between power and growth, and this is something that I didn't adequately explore in my own analyses of power (Katsafanas 2013). But there are different ways of accounting for the relationship between power and growth. For example, Ian Dunkle has argued that Nietzsche's notion of will to power should be expressed as follows:

"Person, A, expresses WP in the course of some activity, Φ , if:

Maximize: A confronts and overcomes more resistance when Φ -ing than A would when not Φ -ing;

or

Grow: A's Φ -ing marks an upward trend in A's confrontation and overcoming of resistance over time" (Dunkle 2020: 205).

As Dunkle understands it, will to power can be expressed *either* in overcoming resistance *or* in growth. These are, effectively, synchronic and diachronic manifestations of will to power. Synchronically, to express will to power is to be in the process of overcoming resistance; diachronically, to express will to power is to manifest upward trajectories in one's capacity to overcome resistance. But these resistances can take various forms, so the trends aren't limited to incorporation. They can also involve growth in ability (as when I get better at some task).

As these remarks indicate, there are important questions about the connection between growth and power. Nietzsche might think that growth is a *necessary* condition for manifesting power, so that someone who remains static in their capacity for overcoming resistance does not count as manifesting power. Or he might think that growth is a *typical result* of willing power, in the sense that as you repeatedly overcome resistances of a certain type, you tend to become capable of overcoming greater degrees of those resistances (for example, as I repeatedly run, I typically become capable of running greater and greater distances). Or he might think that growth is a *normatively privileged* manifestation of will to power, in the sense that while individuals can will power without growing, this somehow counts as defective or even pathological. In light of these alternative possibilities, I wasn't convinced by Richardson's claim that willing power *just is* growing in control by incorporating other wills (57).

But let's move on. Regardless of how power and growth relate, Richardson wants to read Nietzsche as grounding his claims about what we should value on will to power. And he wants to do so in a way that is consistent with Nietzsche's internalism. Richardson thinks Nietzsche is trying to show us that "you should value V because (although you've failed to see this) you already do" (66). But, he notes, this argument "is an effort to persuade you to value what you as yet don't. So it must rely on some distinction between different *parts* of you that do and don't value V or else between different *ways* you do and don't value it" (66). So Nietzsche needs to show either that every person has some *part* that values power or that there is some *way* in which all of us value power.

That task doesn't seem especially daunting. Nietzsche's psychological diagnoses and his genealogies do give us good reasons for believing that even those who reflectively disavow power are, at least in certain respects, willing power. The *Genealogy* is rife with examples, as when Tertullian condemns the quest for power while simultaneously engaging in bloodthirsty revenge fantasies (GM I:15). So let's grant that Nietzsche can establish that there is at least some part of each person that can be described as willing power.

Of course, merely establishing that *one* of my aims is power does not yet establish that I have any particular reason to care about that aim, much less to prioritize it over others. There is part of me that aims at being in Bora Bora, but I've never done a thing to fulfill that aim. So, in addition to establishing that part of each person wills power, we would have to provide some grounds for prioritizing that part over other parts.

Here, Richardson adopts what is sometimes called a True Self view. True Self views hold that there's some constituent, some subpart of me that's *really* me, and that I should prioritize this part over others. As Richardson puts it, Nietzsche's view is that we "should obey.. a principal drive... What 'I'—my agential self, with its worded values—need to learn



is to obey *my genuine self*, which is my body and its great will” (67-8, emphasis added). Thus, Richardson summarizes:

At the bottom of Nietzsche’s argument, I suggest, is a claim as to what’s most truly oneself: what one is. To “internalize” that argument-pattern we need its X to be something truly one’s own; Nietzsche claims that our drives are more truly ‘who we are’ than such authorities as conscience or reason. If we align our agential values to our ruling drive, we are more truly ‘obeying ourselves’. And it’s because his justification cites a valuing that is thus ‘our own’ that it doesn’t need to claim any real or independent value: the value already speaks from us and doesn’t need to speak at us. What’s required is ‘simply’ to notice it. (68)

So Richardson thinks that I can locate something that is most truly me, my genuine self. This will be my drives. And not just any drives, but my ruling drives. So, schematically: I notice that my ruling drive wills power, and that this ruling drive is most truly me; and I thus prioritize it over other aims.

But I think this is questionable, both as a philosophical argument and as an interpretation of Nietzsche. The problem is this: talk of “principal,” “dominant,” or “ruling” drives is ambiguous. For “ruling” can be understood either descriptively or normatively. Descriptively, “ruling drive” would pick out whatever drive *is in fact* motivationally dominant. Normatively, “ruling drive” would pick out whatever drive *should be* motivationally dominant. If Richardson were appealing to the normative sense, his argument would be circular: he would be saying that we should obey drive X because drive X should be obeyed. That’s hardly informative! So I take it that Richardson means “ruling drive” to be read descriptively. He is thus saying that we should obey drive X because drive X is motivationally dominant.

But if we read “ruling drive” descriptively, we seem to run into problems. Consider a case like this: a religious ascetic’s dominant, ruling drive, which organizes and sublimates all of his other drives, is an aggressive drive turned inward. This individual is driven toward all of the activities that Nietzsche mentions in the *Genealogy*: renunciation, rejection of flourishing life, rejection of health, orientation toward an otherworldly goal, and so on. Insofar as the aggressive drive, oriented at the ascetic ideal, is his ruling drive, Richardson’s model seems to entail that he has reason to obey it. He has reason to be as ascetic as possible, to negate physical flourishing as much as possible, to orient himself entirely toward the otherworldly. But this seems problematic.

Now, there would be two ways of resisting this conclusion. One way would be to try to show that the case I’ve just described is impossible: we can’t really have people whose ruling drives are directed toward ascetic renunciation. Your True Self can’t really aim at renunciation. But I find this *very* hard to believe. Are we supposed to believe that individuals such as John the Baptist or Saint Simeon are actually aiming at physical flourishing? That seems rather unlikely. So if we adopt Richardson’s interpretation and accept my claim that people’s ruling drives really can aim at asceticism, then Nietzsche would have no grounds for criticizing these individuals.

The second way of resisting this conclusion would be to complicate the story. Instead of holding that what the agent has reason to do is whatever promotes her ruling drive, we

would maintain that certain ruling drives can block or inhibit some more basic feature. Consider this: perhaps what I have reason to do is *not* whatever the dominant drive aims at; perhaps what I have reason to do is to promote will to power. So, if the dominant drive aims at X, but X is ultimately power-reducing, this would give me reason to abandon the aim of my dominant drive and pursue something else instead.

That second possibility is how I read Nietzsche's ethical theory. The problem with the ascetic priest is *not* that he doesn't follow his dominant drive. He may very well be following his dominant drive. The problem with the ascetic priest is that his dominant drive undermines will to power. His dominant drive disposes him to engage in activities that undermine his long-term capacity either to will power at all or to will power in ways that constitute growth.

Now, I'm somewhat unsure whether Richardson would agree with this. Richardson does say that "there is a bottom to my system of motives and, at this bottom, an aim that sustains all my further motives," namely will to power (70). This sounds somewhat like the second possibility that I've just described. But there are two problems. First, this interpretation seems to me to go against the letter of Richardson's view. For we would no longer say that the individual should obey her ruling drive; we would instead say that the individual should strive for growth in power, and this will sometimes involve *disobeying* or even *rejecting* her ruling drive.

Second, we would have to ask *why* we should strive for will to power. Or, as Richardson puts it, we would have to ask what *authority* will to power has. And I've just suggested that "your dominant drive wills power" isn't a good answer to this question. But I also don't think "will to power is your deepest tendency" is a good answer. To see why, suppose I discover that my deepest motivation is toward some end E. This could be a reason to reinforce my motivation toward E. But it could also be a reason to try to change that fact about myself. To use a Nietzschean example, suppose a modern scientist reads GM III and becomes convinced that his deepest motivation isn't pursuing truth, but inflicting suffering upon himself. He sees this in all that he does. Does that give him a reason to go on doing it or to strive for growth in this activity? It seems to me the answer is clear: not at all. In fact, he might decide to reject this tendency, trying to bring about some change in the way that he lives, the goals that he orients himself toward, the projects that he takes on.

Or take a smaller-scale example: suppose you become convinced that your relationship with Francesca, which you took to be an expression of love, is better conceived as jealous, resentful possession. Diagnosing yourself, you see that your deepest motivation in all of your interactions with her is jealousy, rage, possessiveness, and is very far from love. Does this give you a reason to strive for greater jealousy, more intense rage, deeper possessiveness? Of course not. Sometimes, we have good reason to resist our deepest tendencies. Sometimes, uncovering the fact that X is my deepest tendency should *disgust* me, should spark a desire for change.

Is Nietzsche a constitutivist?

That said, I think Richardson is right that Nietzsche is basically trying to show that we should will power because it is our deepest aim. To make sense of this, though, we need



to say more; we need to offer some argument that shows why we should prioritize will to power over other features of our psychologies.

I think such an argument is available. It's this: will to power is a constitutive feature of human agency. Richardson notes that we could read Nietzsche as a constitutivist who holds that "in being an agent one must value power" (33). Although Richardson says that he "agree[s] that Nietzsche does sometimes try out these arguments," he claims that Nietzsche ultimately embraces some other argument: "I think he opts against this attempt to show that we *must* value power. He prefers to rest his appeal on the claim that we *do*" (34).

But I think this is a potential misunderstanding. As I see it, a constitutivist argument does exactly what Richardson suggests in the last sentence: it shows that we *do* value power. Constitutivism moves from the fact that we *already aim at X* and the claim that *we can't stop aiming at X* to the claim that we *should aim at X*. The basic idea is that insofar as you are acting, you are already aiming at power. Moreover, every episode of action will express this aim, albeit sometimes in a distorted, pathological, or otherwise problematic form. So there's no escaping the aim of power; the question can only be whether you express that aim in a conflicted, attenuated form or an unconflicted, maximal form.

So I think Richardson is underappreciating just how close constitutivism is to the position he advocates and how well it would solve the problems he wants to solve. Let's look at this in more detail. In Katsafanas 2013, I argued that Nietzsche's will to power theory is based on the following claims:

- a) Every episode of willing manifests will to power.
- b) An agent wills power iff she aims to encounter and overcome resistance in the course of pursuing some end.

Richardson seems to accept (a), though he rejects (b) for reasons that I've described above: he thinks power involves growth in control. But let's address this simply by modifying (b) so that it accords with Richardson's reading of will to power:

- (b') An agent wills power iff she aims at growth in control

So far, (a) and (b) are just descriptive claims. We could get normative content out of them by assuming something like this:

If an agent aims at X, the agent has reason to engage in activities that realize or promote X.

For example, if an agent aims at eating, she has reason to go to a restaurant and order food, for those activities are ways of realizing her goal of eating. Or, more to the point, if an agent aims at power, she has reason to engage in activities that realize or promote or manifest power.

Claim (c) is controversial; some philosophers think that aims don't generate reasons. But sentimental theories, neo-Humean theories, and more generally *internalist* theory of practical reason accept versions of it. Richardson clearly accepts some version of (c), though as I pointed out in Section One I am unsure exactly how he would formulate this claim.

So far, it seems to me that we could interpret Richardson's theory in just this way. I think Richardson accepts (a); he certainly accepts (b') and some version of (c). And, if we accept these three claims, we can conclude that in every episode of willing, the agent has reason to engage in activities that realize or promote growth in control (i.e., will to power). That's the conclusion that Richardson wants, and constitutivism gets it. So why isn't Richardson inclined to accept constitutivism?

As I see it, constitutivism differs from standard internalism only in that it includes some claim about what's constitutive of agency. So, if we're trying to decide whether Nietzsche is a constitutivist or a standard internalist, we would just need to determine whether he thinks that power is a constitutive feature of agency. What would that mean? There are different ways of understanding the point, but I hold that X is constitutive of Y iff every instance of Y contains X and containing X is part of what makes Y what it is. So, to claim that power is constitutive of willing, we'd have to show that (1) every instance of willing aims at power and that (2) aiming at power is part of what makes something an episode of willing. There is room for disagreement on whether Nietzsche accepts these claims. It seems to me fairly easy to show that he accepts (1). Claim (2) is harder, as Nietzsche doesn't explicitly put the claim in that way. But he does say that power is the essence of life (GM II.12, GS 349), that power is present in all cases of willing (KSA 13.11[96], KSA 12.2[88]), and so on. So I think there are grounds for reading him in that way, though I won't reconstruct the case here (see Katsafanas 2013 for the details).

Richardson says that he rejects this constitutivist reading because it gives a "transcendental" justification for power, and "this makes the justification more 'coercive' than Nietzsche means it to be... It still assigns to him an externalist justification for his values: they're commanded to us by 'the rules of the game' of agency" (69). I agree that *if* this were what constitutivism did, Nietzsche wouldn't accept it. And I agree that this is how some philosophers, such as David Enoch (2006) interpret constitutivism. But this isn't how I understand constitutivism. Constitutivism isn't a claim about rules of the game. It's a claim about what, in virtue of participating in some activity, we are already doing. In virtue of willing, we're already seeking power. The constitutivist argument shows us both that we're already doing this and that it's not contingent, that we can't stop doing it.

Here's one way of thinking about this. A *transcendental* argument operates by showing that something is a condition for the possibility of something else. For example, Kant tries to show that the Categories are conditions for the possibility of self-conscious experience. We know that we have self-conscious experience; so, if the arguments about the Categories are correct, we are entitled to conclude that the Categories apply to every episode of conscious experience. We can distinguish this from a mere claim about regularity: it might be that, as a matter of empirical fact, most or even all human conscious experiences involve (let's say) three spatial dimensions; but perhaps there could be some conscious creature that experienced things in only two spatial dimensions.

Richardson is assuming that my constitutivist reconstruction of Nietzsche takes a transcendental form. I don't myself view it that way. As I see it, Nietzsche is arguing that as a matter of empirical fact all human agency is motivated by drives; and he is claiming that part of what it is for something to be motivated by drives is for it to exhibit will to power. The latter claim is definitional or, if you like, analytic: drives are simply *defined* as dispositions toward continuous manifestation of some form of activity (which I take to be



equivalent to the claim that drives are simply defined as dispositions that exhibit will to power; see Katsafanas 2013 for an explanation). I *don't* think that Nietzsche is claiming that willing power is a condition for the possibility of agency. There could be forms of agency that didn't exhibit will to power. But they would be radically different forms of agency, unlike the forms of agency exhibited by drive-motivated creatures. Analogously: there could be forms of conscious perception that are merely two-dimensional, but they would be radically different from human forms of conscious perception.

So, as I see it, the claim about constitutive features just establishes the universality and immutability of the relevant feature. It doesn't establish that the feature is a transcendental condition for all possible versions of agency.²

If we accept this constitutivist reading of Nietzsche, we would have an answer to the question “why should we value power” and “why should power have priority over other things we value.” Constitutivism can answer these questions in a way that standard internalism cannot. For constitutivism locates a feature that is both omnipresent and unchangeable, a feature that is a fixed part of agency or of willing and hence can't be avoided. You can try change particular motivations, dispositions, values, and habits, but there is no way, even in principle, of making it the case that your striving no longer manifests will power. And yet certain ways of striving—certain configurations of affect, custom, disposition, habit, and value—generate conflicts with will to power. Assuming that we have some reason for avoiding these conflicts, we would have reason to shed those aspects of ourselves that generate the conflicts. Put differently, we have reason to modify our values, habits, dispositions, customs, and affects so that they minimize conflicts with will to power.

To summarize, then: constitutivism just shows that will to power is an unchangeable and omnipresent feature of drive-motivated action. Just as the biology of our visual system commits us to seeing in three dimensions, so too the biology of our drives commits us to having a background aim of power in addition to any other particular aims that we might have. We can try to revolt against this, but even our revolts will end up re-expressing will to power. So our choices are to express will to power in an unconflicted form or to express it in a conflicted form.³

The distinction between animal drives and socially induced drives

I will close with a somewhat different point. On 191, Richardson distinguishes between an original layer of animal drives and a new layer of socially-induced drives. He explains that

Nietzsche's principal idea is that agency was selected for the good of the social group. It is a product of the prehistoric work to 'tame' the human animal... This installs agency as a second system of dispositions in us, one superimposed on the

² On pages 175-7, Richardson argues that Nietzsche transforms Kant's transcendental arguments into claims about the conditions of life. I found this section fascinating. I think that's one way of describing how I see Nietzschean constitutivism: it shows that power is a condition of our form of life.

³ Of course, we can choose conflict and disorder. Nothing prevents us from doing so; nothing prevents us from turning on ourselves. This is why I argued that a notion of affirmation or endorsement also plays a crucial role in the Nietzschean version of constitutivism. In short: Nietzsche needs to appeal to two different constitutive aims, one involving power and the other involving affirmation. See Katsafanas 2013 for the details.

original 'animal' drives and affects. So there is a deep conflict between agency and those original drives. (189)

The idea, then, is that pre-societal human beings would have drives that work for the individual's own good, whereas societal human beings have a new layer of motivations that work for the good of the social group. Richardson summarizes:

Nietzsche's drive psychology works with a broad distinction between (a) drives rooted in the 'animal' past of human and counted as 'body' and (b) drives enabling the distinctive human practices of consciousness and language and counted as 'spirit'. Nietzsche claims that these latter drives were *imposed* on the first set by social pressures, not designed for their own betterment. (191)

I found this unconvincing. I think this dichotomy is too stark. There are philosophers who operate with this kind of dichotomy between the pre-societal human and the socialized human. Rousseau is an example: the harmonious inner life of the pre-societal person is spoiled by his socialization, distorted by the pressures of amour propre, which leads us to dissimulation and corruption. But I think for Nietzsche this is the wrong way around. The individual is a late outgrowth rather than a starting point. Consider *Daybreak* 105. The passage begins like this:

Whatever they may think and say about their 'egoism', the great majority nonetheless do nothing for their ego their whole life long: what they do is done for the phantom of their ego which has formed itself in the heads of those around them and has been communicated to them; as a consequence they all of them dwell in a fog of impersonal, semi-personal opinions, and arbitrary, as it were poetic evaluations, the one for ever in the head of someone else, and the head of this someone else again in the heads of others: a strange world of phantasms which at the same time knows how to put on so sober an appearance!

In those lines, Nietzsche sounds like La Rochefoucauld or Rousseau. That is, it sounds like Nietzsche is saying that our attempts to act in our own interest are distorted by our interactions with others. We strive to make others think well of us; accordingly, we are driven to dissimulate to others who are also dissimulating to us, so that all of social life is like a fog of vain striving. That *does* sound like it fits with Richardson's reading: if we could just break out of this social setting, we could ignore the phantoms of public opinion and be true to our selves. We wouldn't have to care about how others see us; we could fall back to free, undistorted expression of our pre-social selves.

But that is *not* how Nietzsche concludes the section. Here is the rest of the passage:

This fog of habits and opinions lives and grows almost independently of the people it envelops; it is this fog that their lies the tremendous effect of general judgments about 'man' – all these people, unknown to themselves, believe in the bloodless abstraction 'man', that is to say, in a fiction; and every alteration effected to this abstraction by the judgments of individual powerful figures (such as priests and philosophers) produces an extraordinary and grossly disproportionate effect on the great majority—*all because no individual among this majority is capable of setting up a real ego, accessible to him and fathomed by him*, in opposition to the general pale fiction and thereby annihilating it. (D 105, italics added)



Consider Nietzsche's explanation, given in the italicized lines. The Rousseau-ian picture is this: we dissimulate in order to achieve social esteem, thereby hiding and perhaps even corrupting our true selves. The Nietzschean picture is very different: we don't even have true selves to hide. The only selves we have are the fantastical ones, the ones constructed of social imagery and opinion. There is no undistorted true self lying behind the social manifestations of the self; there is *only* the one layer, the socialized impersonal self. To move past this is possible. But to move past it would not consist in freeing a trapped self from the accretions of culture. To move past it would be to create something new. This would have to be "set up", created and "fathomed" by the person, not merely found. And this is no isolated idea in Nietzsche: "Truly high respect one can have only for those who do not *seek* themselves" (BGE 266).

So I think Nietzsche rejects the idea that there is a pre-social set of bodily drives, aiming at individual flourishing, which is somehow corrupted by a social level of drives that aim at group flourishing. His picture is much more complicated.

But I don't know if this is really a criticism of Richardson, for in a way he seems to agree with some of these points. After all, as he explains in Chapter 10, he thinks that Nietzsche is urging us to form a new kind of self, which sees and incorporates the truth about its own drives, affects, and perspectives. We have to "free ourselves from forces—in drives, in society—that had always covertly controlled us" (415), and this not only "make[s] possible a more genuine freedom than the kind we long supposed we had," but also "advances our *selfhood*. In recognizing and then freeing ourselves from foreign or alien forces, we take authority for ourselves and 'become who we are'" (415). And this new self will take up two tasks: "it will strive to make itself more genuinely *one*, out of its drives, and also to become more genuinely *its own*, our of its group" (417). Put differently, it will struggle to fashion some kind of unity instead of fragmentation of drives; and it will try to assimilate social identities into something individual (417). The latter point is explained as follows: "one is not well differentiated from the groups to which one belongs, one merges into a kind of group self in them," so that the "challenge in making a self is to crystalize it out of the social matrix in which it is originally set" (424). To do this, I have to "make my values my own," by revaluing them: "I must *make it the case* that my values express my own perspective" (431). To do so is to express a "genuine" form of selfishness: "to make my selfishness genuine I need to recognize the other forces that have dictated my values to me," where these other forces range from my drives, to social constraints, to normative claims about virtues and vices, to social expectations, and so on (436). In so doing, one potentially founds new social norms, which Richardson discusses in Chapter 11.

All of this seems exactly right to me. And it's also emblematic of the way that Richardson's text operates. Just when you think he's missed some aspect of Nietzsche, it turns out that he circles back to it after all.

Works Cited

- Dunkle, Ian. "On the Normativity of Nietzsche's Will to Power." *The Journal of Nietzsche Studies* 51:2 (2020): 188-211.
- Enoch, David. "Agency, Shmagency: Why Normativity Won't Come from What is Constitutive of Action." *Philosophical Review* 115:2 (2006): 169-198.
- Katsafanas, Paul. *Agency and the Foundations of Ethics: Nietzschean Constitutivism*. Oxford University Press, 2013.

154 *Review of John Richardson, Nietzsche's Values*

Reginster, Bernard. *The Affirmation of Life*. Harvard University Press, 2006.

Richardson, John. *Nietzsche's Values*. Oxford University Press, 2020.

