Recent Work on Nietzsche's Moral Psychology and Ethics

Paul Katsafanas

Forthcoming in *Nietzsche-Studien*

Penultimate draft—please cite/quote the published version

Alfano: *Nietzsche's Moral Psychology* (Cambridge University Press 2019).
Leiter: *Moral Psychology with Nietzsche* (Oxford University Press 2019)
Ridley: *The Deed is Everything: Nietzsche on Will and Action* (Oxford University Press 2018)
Stern: *Nietzsche's Ethics* (Cambridge University Press 2020)

These four books are broadly on Nietzsche's moral psychology and ethics. The books differ widely in their aspirations: Ridley's is focused solely on Nietzsche's notion of action, whereas Leiter's is more synoptic. And they also differ widely in their conclusions: Leiter presents Nietzsche as a nearly infallible figure who has not only solved longstanding philosophical problems but has even managed to anticipate recent results in empirical psychology. Stern, on the other hand, presents Nietzsche as a rather amateurish philosopher, who picks up dribs and drabs from his cultural context and amalgamates them into interesting and provocative, but indefensible, positions. Between these extremes, we have readers like Ridley and Alfano, who are not averse to pointing out lacunae in Nietzsche's arguments but who nonetheless see him as deeply insightful.

Although each of these books is worth reading, I will argue that they have various degrees of success. Alfano's book is, to my mind, the most successful at achieving its stated aims; while I point out some potential oversights and some areas that could benefit from further development, Alfano's book is both novel and important. Leiter's book is clearly written and presents the arguments in an admirably forthright manner, but some of its conclusions are vitiated by lapses and mischaracterizations. Stern gets Nietzsche's basic view right, but does not probe it very deeply and is too quick to present Nietzsche as confused; I see the confusions as emanating less from Nietzsche's texts and more from Stern's reading of them. Ridley's book is original and provocative, but I find the central claim—that Nietzsche endorses an expressive account of action—ultimately unconvincing. Nonetheless, even the books I regard as flawed are valuable, for reasons I will point out along the way.

~ ~ ~

I will start with Tom Stern's *Nietzsche's Ethics,* which is both engaging and frustrating. Stern is a good writer, with a gift for pithy formulations, and this makes his book an enjoyable read. However, when it comes to the articulation of philosophical positions, Stern tends to stop at the surface: he ends up attributing to Nietzsche views that are strikingly underdeveloped.

Stern's goal is to explain Nietzsche's ethical theory as it emerges in his works from 1886 onwards. Stern takes Nietzsche's ethical theory to be quite simple: life has certain aims; it is our job to carry out these aims; when we fail to do so, we are ethically criticizable. Let me explain.

The book begins with a discussion of what Stern calls the "Life Theory." The Life Theory states that organisms seek dominance and control; in brief, organisms seek power. Moreover, this aim is "biologically essential" for living organisms: "*living* and *power seeking* cannot be pulled apart" (6). So, "when Nietzsche looks at the realm of living things, what he sees is a domain necessarily characterized by power seeking" (6).

The Life Theory is a descriptive thesis which is "presupposed by Nietzsche's ethics" (7). So what is that ethics? Well, "Nietzsche's basic ethical position is as follows: *it is ethical to further the goals of Life, and it is unethical to impede them*" (11). Or again, "Nietzsche's normative command is: Affirm Life!" (13). Importantly, Stern thinks that affirming life requires nothing more than pursuing power: "affirmation means to increase, seek power, explain, exploit, while Life-denial means the opposite" (12). In sum, then, Stern reads Nietzsche as having a very straightforward ethical theory: we should express power as best we can (11-15).

At a high level of generality, I think this is basically correct as a reading of Nietzsche. But it does leave many unanswered questions. What is power, exactly? What does it mean to pursue power? What is it to further the goals of Life—is this simply to maximize power, or is there some more complicated set of criteria? Stern's book contains remarkably little on these sorts of questions; he does not develop these generalized claims into anything more specific.[1] So that is one complaint about the book: although one obviously cannot elaborate every claim in full detail, a book that bills itself as providing an explanation of Nietzsche's ethical theory should at address at least some of these questions, or, barring that, should at a bare minimum flag them as things that stand in need of further explanation.

This brings me to an important methodological point. Stern is happy to distinguish the question of what Nietzsche believes from the question of what seems philosophically defensible to us. Thus, Stern writes that although Nietzsche endorses the Life Theory, "This does not mean that the Life Theory grounds Nietzsche's ethics *unproblematically;* indeed, my analysis will suggest the opposite… Moreover… the Life Theory need not amount to the interpretation that is most complex or satisfying to the modern reader" (10-11). Or, even stronger: Nietzsche "has to say *why* we ought to further Life's goals and, I will suggest, he has no good answer" (29). In short, Stern thinks that Nietzsche lacks any good arguments both for the Life Theory itself (i.e., the view that power-seeking is biologically essential) and for the move from the Life Theory to the ethical view.[2]

Why is this? Oddly, Stern claims that his reading of Nietzsche's ethical theory bears a "superficial resemblance" to my own reading of Nietzsche. Perhaps because he only cites an article that describes my view in passing (Katsafanas 2018) rather than my book on the topic (Katsafanas 2013), Stern does not see that his reading of Nietzsche's ethics is (at a high level of generality) identical to the reading of Nietzsche that I defend; the key difference lies only in that I think Nietzsche has a good, agency-based argument for the claim that life or power has a privileged normative standing. So, at a high level of generality, Stern and I agree on the manner in which Nietzsche critiques values,

---

[1] Stern says a bit about power on pp. 15-29, in the context of discussing Nietzsche's critique of Christianity; there, he points out that Nietzsche associates power with exploitation, helping the weak to perish, promoting military efforts, avoiding certain kinds of metaphysical and religious views, avoiding world-denial, and so forth. But he does not move from this list of things associated with power to an analysis of what power is; in fact, he suggests that there is not enough material in the texts to do so.

[2] Stern does think that Nietzsche has some biological evidence for the Life Theory, drawn from Rolph, Roux, and others.

persons, and cultures: Nietzsche employs a descriptive principle (namely that life in some sense aims at power) in order to justify his ethical claims. Stern and I disagree only on how Nietzsche attempts to justify this method of critique. For Stern, there is an ambiguously stated biological claim about power-seeking and an (ultimately indefensible) claim that we are ethically committed to power-seeking; for me, there is a more precise claim about power-seeking that Nietzsche defends via his drive psychology and then employs in order to support his normative claims about our commitment to power-seeking. In other words: I think Nietzsche has good arguments for his claims about the normative standing of power, whereas Stern concludes that he does not. Of course, my interpretation of Nietzsche's arguments might be wrong; I certainly do not pretend to have established my reading with apodictic certainty. And in a short book like Stern's, an author obviously cannot address every possible interpretation. So my complaint is not that Stern neglects my view in particular. My complaint is that before concluding that Nietzsche's ethical theory is fatally ambiguous and argumentatively ungrounded, a book on Nietzsche's ethics should consider at least one or two closely related readings that present Nietzsche as actually having a good argument. (I have mentioned only my own view, but readings by Bernard Reginster and John Richardson would be relevant as well; they, too, see Nietzsche as successfully linking descriptive claims about power to normative claims.)

But let us move on. Having presented the Life Theory, Stern turns to its critical function. He focuses on how the Life Theory can be used to critique Christian morality (Section 2). Stern argues that Nietzsche has two different strategies, one of which relies on the idea that Christian morality "prevents Life from achieving its goals" (26) and the other of which claims that while Christian morality purports to be anti-Life, it actually promotes Life (26). This seems to me entirely correct, although I was surprised to see that Stern devotes a lengthy section to arguing that these two strategies are in tension with one another (pp. 25-30). As he himself remarks on p. 27 they need not be: Nietzsche's idea is not that Christianity totally prevents or fully promotes Life, but rather that it undermines Life; Christianity makes it less likely that Life will operate as best it could (27).

This is as good a point as any to bring up something that bothered me about Stern's book: Stern will often claim that Nietzsche faces some difficult, irresolvable philosophical problem, when in fact the problem dissolves with a moment's analysis. Thus, despite spending five pages (or 8% of this 61 page book!) arguing that there is a dramatic and insuperable difference between the claim *Christianity prevents Life from achieving its goals* and *Christianity promotes Life's goals,* it is just obvious that the first claim would be in tension with the second only if we read the first claim uncharitably. By analogy: if I say "heroin prevents you from achieving happiness" in one breath and "heroin promotes happiness" in the next breath, there need be no contradiction; the statements can be understood as simply contrasting total and partial fulfillment of some aim. Just as heroin might generate some immediate happiness while undermining one's capacity to enjoy a more complete or more lasting form of happiness, so too Christianity might realize Life's aims to some extent in some circumstances while undermining one's capacity for a more complete or more lasting fulfillment of Life's aims.[3]

---

[3] Stern sometimes states the second claim as follows: Christian morality "conceals a strategy on Life's part to achieve its goals *as best it can*" (26, emphasis added). Or, a bit later: Christians manifest will to power "to the maximum of their ability" (28). Thus, we might interpret Stern as claiming that there is a tension in Nietzsche between (a) the claim that Christianity prevents Life from achieving its goals, and (b) the claim that Christianity *maximally realizes* Life's goals. If this is what he intends, though, Stern seems to me to be misinterpreting Nietzsche; Nietzsche nowhere endorses (b). Stern does cite some passages in which Nietzsche claims that *particular individuals* maximally manifest Life's aims by promoting Christian morality; but we cannot move from the claim that (say) St. Paul's best way of maximizing Life's aims is by

This is just one example, but I found myself struck by similar thoughts throughout Stern's book. I certainly respect Stern's desire to avoid presenting Nietzsche as infallible, but Stern goes *very* far indeed in the other direction. Stern presents Nietzsche as confused or ambivalent about most of the philosophical issues he discusses, but the alleged confusions and ambivalences are either very superficial or depend on highly uncharitable readings of the texts. You could, for example, accuse me of being confused if I utter those two statements about heroin and happiness; but that would be priggish. So, why spend so much time presenting (rather superficial) tensions or ambivalences, and so little time thinking about how they might be resolved or might point to some deeper view? I was. not sure. And that question is especially pressing given that Stern bills this book as an introduction to Nietzsche. Do newcomers to Nietzsche really need to be told that there are apparent tensions and contradictions in his works? That is obvious. What is not obvious, what takes more philosophical work, is uncovering the actual views that Nietzsche is trying to articulate.

Part of the problem, here, is that Stern's methodology is driven by exaggerated dichotomies. He suggests that

> a great deal of philosophical writing about Nietzsche… is creative in method and aim. It aims to produce the 'best' Nietzsche… the one most attractive to present-day Anglophone philosophers – and it is willing to do so at the expense of what its practitioners might see as an inflexible … preoccupation with fidelity to the texts and their historical context… My own intention is to stay very close to the texts … and to present the philosophical ideas found in them as clearly, neutrally, and thoroughly as possible. (3)

The suggestion, here and elsewhere, is that Stern is unique among commentators in aiming for textual fidelity: whereas other commentators supplement and distort Nietzsche's arguments in order to produce a more appealing Nietzsche, Stern lets Nietzsche's own claims shine through, unfiltered and cleared of interpretive additions.

In a way, that is a respectable goal. It is certainly true that some commentators supplement Nietzsche's claims with their own arguments; it is also true that, at times, commentators fail to distinguish moments at which they are adding to the texts from moments at which they are interpreting Nietzsche's own claims. But it is wrong to suggest that there are only two options, pointing to what the texts actually says or adding in arguments. The truth is that there is a spectrum. At one extreme, you would merely have a list of quotations from Nietzsche's texts, presented without commentary; at the other, you would have a philosopher who is very loosely inspired by Nietzsche's ideas but is inventing her own arguments. In between, you have varying degrees of interpretive, reconstructive, and elaborative work. Nietzsche might phrase something ambiguously or unclearly, and the interpreter might look to other passages to clarify what Nietzsche means. Nietzsche might invoke some distinction that we can draw more clearly. Nietzsche's arguments might have suppressed premises that he does not explicitly flag, but which we have good reason to think he would endorse. Nietzsche might leave something implicit that the commentator chooses to make explicit. Nietzsche might make a hazy or elliptical reference to something that the commentator states more directly. And so on. Stern's approach is not, as he suggests, to simply be a finger that points to what Nietzsche actually said. On the contrary, Stern's presentation is chock

---

being a Christian to the claim that *in general* Christianity maximizes Life's aims. Consider an analogy: it might be true that for particular unfortunate individuals, taking heroin is the best way of maximizing their happiness; but it hardly follows that *in general* taking heroin is the best way of maximizing happiness.

full of contestable interpretations. Consider, for example, Stern's insistence that Nietzsche's Life Theory is biological rather than being based on claims about "action and drives" (10). This looks to me like an imposition on the text, and a faulty one at that. Stern's seems to be assuming that claims about drives and action are not biological claims; I suppose, instead, that he takes them to be psychological. But Nietzsche does not operate with a biology/psychology distinction and, were he to draw that distinction, it is highly unlikely that he would draw it in the way we do today. So what sense can be given to the question of whether *drives* are biological or psychological? Or, for another example: why should we assume that when Nietzsche talks about Christianity realizing Life's aims, he means that Christianity maximally realizes life's aims at all times and in all circumstances, instead of just reading him as making a localized point about how certain ascetic priests in a certain historical period manage to maximize their own expression of power by promulgating Christian views? As these examples indicate, Stern's presentations of Nietzsche's claims rely on interpretive choices, which are not forced on us by the text. Stern is right that we should be wary of excessively charitable interpretations (he has a lovely critique of some interpretations in Stern 2018, as well as an illuminating discussion of the principle of charity in Stern 2016; I highly recommend both of these pieces). But the answer is not to flip to the other extreme, pretending that one's own presentations of Nietzsche's views are unique in giving us the uninterpreted real Nietzsche. On the contrary, the solution is to assess alternative interpretations piecemeal, sorting out what is adequate and what is inadequate, what is supplementary and what is there in the texts. And this is what, I am suggesting, Stern fails to do.

But let us move on. The following three sections turn to the role of genealogy. Stern briefly canvasses three roles that genealogical investigations can play: to provide evidence for the Life Theory; to show us how we ended up with our current moral system; and to persuade us to adopt a Life-Affirming ethic. He offers a reading of the *Genealogy* which emphasizes these three features; and then he asks, in Section 6, whether the *Genealogy* succeeds at these goals. It comes as little surprise that Stern's answer is a tentative 'no': the evidence for the Life Theory is spotty and questionable (46), the story of the transition from noble to slave morality does not have a good answer as to why the nobles are susceptible to the slave morality (48-51), and (Stern claims) the best argument for the idea that we should affirm life rests on saying that otherwise we will be "divided or alienated" (53). So again, Nietzsche is presented as coming up short.

Again, I wished for more. Take the third point. Suppose it is true that Nietzsche's best argument for life-affirmation comes down to the idea that we have two alternatives: either affirm life, in which case you will be unified; or do not, in which case you will be disunified. (I think this grossly misrepresents the complexity of Nietzsche's thoughts on affirmation, but set that aside.) Stern finds this dissatisfying because, he claims, we might simply prefer to be disunified. Now, Stern is certainly right that we can ask *why* we should prefer unity to disunity; and he is right that a divided and alienated existence can look appealing or even "tragic-heroic" (53). In short, it is true that we can ask "is anything especially wrong with being so divided?"(53). But Nietzsche—and with him a whole tradition of 18th and 19th century philosophers, from Rousseau to Hölderlin to Goethe to Schiller to Hegel to Marx, have considered and addressed this question. To put this in Stern-ian terminology: there is a Team Unity, who argue against Team Disunity. It is one thing to assess the arguments; perhaps they are not decisive, and in any event the notions of Unity and Disunity are multifarious. But it is clear that Nietzsche is on Team Unity. So, when we pose Stern's question—is anything wrong with exhibiting Disunity?—Nietzsche will say *yes*. And then he will tell you what, in particular, is wrong. But, reading Stern, you would not know this.

So, to sum up: this book does a nice job of presenting the gist of Nietzsche's views in a straightforward fashion. And in one respect I quite liked the uncharitable readings of Nietzsche; it is refreshing to see a Nietzsche scholar admit that Nietzsche can be wrong, that Nietzsche has underdeveloped views on some points, that his arguments are not always worked out, that his uptake of certain ideas is rather spotty. But Stern seems to delight in presenting Nietzsche as incapable of defending his views and as unaware of obvious inconsistencies. Nietzsche is far from perfect, but he is a much better philosopher than Stern recognizes.

~ ~ ~

Brian Leiter's *Moral Psychology with Nietzsche* is at the opposite end of the spectrum: whereas Stern presents Nietzsche as largely unsuccessful, Leiter portrays Nietzsche as having (what Leiter regards as) the best possible philosophical views and even as having anticipated claims in recent empirical psychology.

Leiter's book is based almost entirely on previously published papers. Although there are some light revisions and expansions, the key claims and arguments will be familiar to anyone who has read Leiter's previous work. In particular, Leiter takes Nietzsche to endorse the following six claims: (1) epiphenomenalism about consciousness; (2) denial of freedom; (3) anti-realism about moral value; (4) anti-realism about epistemic value; (5) denial of "privilege" readings of power; (6) speculative methodological naturalism. I accept (3) and a version of (6), but think there are problems with (1), (2), (4), (5). However, I only have space to address (1) and (5).

We can start with (1). The claim that a mental state is epiphenomenal typically means that the mental state has no causal effects. It sometimes *sounds* like Leiter uses the term in this way. For example, Leiter attributes to Nietzsche the view that "the kinds of self-conscious mental states that sometimes precede action are epiphenomenal" (138). On that reading, it sounds like Nietzsche is saying that self-conscious beliefs, desires, intentions, thoughts about values, and so forth have no causal effects. So, for example, a Christian might form the conscious belief that sex is sinful or that God requires devotion, but these conscious beliefs will have no causal impact on his actions or his other mental states.

But Leiter actually backs off from that claim, instead associating epiphenomenalism with the more modest view that "the causal efficacy of a wide variety of mental states *does not depend on their being conscious*" (138, italics added). In other words, he wants to maintain that the Christian's conscious belief that sex is sinful *can* be causally efficacious, but deny that its being causally efficacious depends on its being conscious. We could put the point this way:

> (Epiphenomenalism) if we consider some mental state M, it will play exactly the same causal role regardless of whether M is conscious or non-conscious.

There are philosophical views according to which Epiphenomenalism is or could be true. For example, the higher-order thought theory of consciousness holds that a mental state M is conscious just in case the agent is aware of being in that state. If we combine the higher-order thought theory of consciousness with the claim that the conscious state is not changed in any way when it becomes conscious, then Epiphenomenalism could be true: the agent's awareness of being in state M could be causally inert, with of all the causal work done by M itself. Leiter, following Mattia Riccardi, is tempted to attribute this view to Nietzsche.

But there is a complication that Leiter does not consider: if mental states are in some way *changed* or *transformed* by becoming conscious (i.e., by becoming an object of the agent's awareness), then Epiphenomenalism (as defined above) could be *true but irrelevant.* Consider: when mental state M becomes conscious, it is no longer the same state as before. So, rather than speaking of one state, M, which abides regardless of whether it is conscious or unconscious, we have to speak of two states, M1 and M2, where M1 is the original non-conscious state and M2 is the conscious state. In Katsafanas 2016, I argue that this is in fact Nietzsche's model of consciousness: Nietzsche claims that conscious states are distorted, generalized, or rendered superficial in the process of becoming conscious.

Let me explain this in a bit more detail. We can contrast two models of the unconscious/conscious relation. On the Single Stream model, we have a single state, M, that can be either unconscious or conscious. At any given moment, the state is either conscious or unconscious. On the Two Stream model, we have two states: an unconscious M1 which gives rise to a conscious M2. M1 could persist even when M2 arises, so at any given moment we might have two states, M1 and M2. Freud is a familiar example of someone who endorses the Two Steam model. For Freud, an unconscious wish can persist alongside its conscious counterpart, a desire. Or an unconscious phantasy can persist alongside the conscious ideal to which it gives rise. In Katsafanas 2016 (Chapter 2), I draw on textual, historical and philosophical arguments to show that Nietzsche endorses a version of the Two Stream model of the unconscious-conscious relation. (I argue that in so doing, he is developing ideas from Carus, Herder, and Hartmann.) So, for example, when Nietzsche discusses the unconscious bad conscience giving rise to conscious guilt, his idea is that an unconscious state persists alongside the conscious states to which it gives rise; his idea is *not* that we have a single state of which we are sometimes aware and sometimes unaware.

Why does this matter? Because it changes the import of Epiphenomenalism. If Nietzsche thinks that states are transformed in the process of becoming conscious, then questions about whether he is an epiphenomenalist become questions about whether the transformed or distorted conscious states are causally efficacious. Obviously they can be: Nietzsche repeatedly asserts that the distortions of consciousness have causal effects (e.g., GS 44, GM II; cf. Katsafanas 2016 Chapter 2 for many more passages). Now, we could reconcile this with Epiphenomenalism by saying that an undistorted M1 becomes a conscious, distorted M2; although M2 is initially consciousness, it can become non-conscious and can continue to produce causal effects regardless of whether it is presently conscious or non-conscious. So Epiphenomenalism (as I define it above) could be true. And that seems fine to me as a reading of Nietzsche. But then notice that Epiphenomenalism does not really engage with Nietzsche's primary concerns. After all, even if *particular* mental states have causal effects independently of being objects of inner awareness, Nietzsche's point is that the *becoming conscious* of mental states has causal effects: it introduces causally consequential distortions into the agent's mental economy. That is why Nietzsche routinely emphasizes the dangers and "sicknesses" that consciousness produces. And that is what my reading of Nietzsche on consciousness addresses. So the energy that Leiter devotes to defending Epiphenomenalism seems to me misplaced; the view that he defends is a bit of a red herring.

Which brings me to another point about Leiter on consciousness. Leiter spends a fair amount of time arguing against a detail of my reading of Nietzsche on consciousness. I argue that Nietzsche associates the conscious with conceptually structured mental states, and the unconscious with states that lack conceptual articulation. Leiter becomes exercised by the question of whether conscious

states have to be conceptual or whether they can be linguistic-but-not-conceptual. I suggested that Nietzsche uses the terms conceptual and linguistic interchangeably (Katsafanas 2016, Chapters 2-3), and thus phrased the conscious/unconscious distinction as the conceptual/nonconceptual distinction: conscious states have conceptual contents, whereas unconscious states have nonconceptually articulated content. Leiter insists that Nietzsche must have recognized a crucial distinction between conceptual states and linguistic states, although Leiter presents no direct textual evidence for this claim; Leiter's attribution of this distinction to Nietzsche is based solely on the claim that we need the distinction in order to avoid certain philosophical problems that Nietzsche himself never mentions (135-138). I do not see why Leiter thinks we should force on Nietzsche a subtle distinction that is not typically drawn in his historical context (see Katsafanas 2016), that Nietzsche nowhere explicitly mentions, and that none of Nietzsche's arguments depend upon. In other words, I think you would need strong textual evidence to establish that Nietzsche *distinguishes* conceptual and linguistic content, whereas Leiter is operating with the (to my mind mistaken) assumption that we need strong textual evidence to show that Nietzsche *identifies* the two. But, regardless of who is right about the burden of proof, I actually do not think that the dispute really matters. What is crucial is the fact that Nietzsche distinguishes conscious and unconscious states primarily in terms of their *structure,* rather than in terms of *inner awareness;* debates about whether this structure should be understood primarily as linguistic or primarily as conceptual are largely inconsequential (in the sense that answers to this question do not affect any of Nietzsche's other claims or arguments).

Let us move on to claim (5). Leiter's Nietzsche is not only an anti-realist about value, but also someone who thinks that there are no grounds for privileging one set of values over others. Although this may not be apparent in Leiter's framing of the issues, it is important to note that these are *distinct* claims. You could accept the first claim while denying the second: you could think (as I do) that Nietzsche is an anti-realist in the sense that he denies that values would exist independently of human subjectivity, while maintaining that facts about our subjectivity nonetheless give us grounds for condemning certain sets of values and praising others. Reading Nietzsche in this way (as an anti-realist who nonetheless thinks that there are facts about which sets of values are acceptable) fits nicely with the texts. After all, Nietzsche repeatedly assesses individuals, cultures, values, and so forth in terms of ascending and descending life; indeed, in the late texts, Nietzsche just straightforwardly asserts that power, health, and/or life are the criteria of evaluation that we should employ (e.g., A 2; for further references, see my comments on Stern, above, and Alfano, below). So what does Leiter do with those passages? Well, if John Stuart Mill says that you should assess actions in terms of utility, most people would read Mill as stating just that: we should assess actions in terms of utility. Or, if Kant says we should assess actions via the Categorical Imperative, most of us would read Kant as meaning just that: we should assess actions in terms of the Categorical Imperative. But Leiter wants to read Nietzsche in a different way: "Nietzsche's own evaluative perspective in undertaking the revaluation [of values] does not, in fact, enjoy any metaphysical or epistemic privilege over its target" (65). So, when Nietzsche says that we should assess values, persons, or cultures in terms of life or health or power, he does not really mean that. Instead, Nietzsche means that he happens to prefer those people who evaluate things in this way, but "at bottom, Nietzsche has nothing to say to those readers who don't share his evaluative tastes" (65). (For someone who has nothing to say, Nietzsche certainly says quite a lot; but let us set that aside.)

This is an absolutely crucial move for Leiter; in order to support the Anti-Privilege reading, Leiter needs to dismiss the literal reading of Nietzsche's claims in favor of something more tentative and

qualified. So how does Leiter respond to those who take Nietzsche's claims about power literally? Rather than considering the careful, analytically precise work on this topic that has emerged in the past decade or so, Leiter examines a fifty-year old paper by Philippa Foot and an offhand, forty-year old remark by Richard Schacht, in which Schacht claims that Nietzsche assesses values in terms of power. Leiter correctly says that in order to evaluate this claim, we "require a more precise statement" of the argument (53). But rather than consider any *actual* interpretations or defenses of this claim that have emerged in the last forty years, Leiter suggests that anyone who reads Nietzsche as giving power a privileged normative status must be embracing some analogue of John Stuart Mill's argument: "to show that something is desirable (i.e., valuable), show that it is desired" (53). That is: Leiter assumes that in order for Nietzsche to believe that power has a privileged normative status, Nietzsche would need to rely on the claim that because power is desired it is desirable. Leiter then spends several pages rehashing standard objections to this kind of claim, reiterating old chestnuts about how we cannot move directly from claims about what is desired to claims about what is desirable. Leiter claims that because these arguments are so bad, Nietzsche could not have believed them; hence, Leiter concludes, Nietzsche could not have really thought that power had a privileged normative status.

But this argument is exceptionally weak. Suppose we grant every step of Leiter's argument. Leiter would then succeed in showing Nietzsche's argument for the normative privileging of power cannot be a version of Mill's argument. But Mill's argument is just one of *many* ways of moving from claims about human nature, motivation, or affects to claims about what is valuable; moreover, it is an argument that Nietzsche nowhere explicitly mentions. So I am left wondering: why not consider what Nietzsche actually says about these matters? Why not look at the texts or the secondary literature and consider some real possibilities?

Unfortunately, Leiter does not do so. Instead, after criticizing the Millian argument, Leiter considers whether Nietzsche might hold that "power is the only thing that is, in fact, desired" (56). Leiter claims "many, of course, have thought that Nietzsche held precisely this view," although he does not cite a single person who actually thinks this (56). He does claim, a few pages later, that I believe that will to power is "the exclusive motivation for all human behavior" (59 fn. 16). That is false: Katsafanas 2013 defends the view that power is a *ubiquitous* human motivation, not that it is the *only* human motivation; and in fact I spend several sections explicitly arguing *against* the possibility of power's being the only human motivation or the only evaluative standard (Katsafanas 2013: 160-164 and 176-197).

So far, we have nothing that supports Leiter's dismissal of the literal reading of Nietzsche's ethical claims. But Leiter does have some additional arguments. He gives three quick considerations against the "strong form of Nietzsche's descriptive doctrine of will to power" (53). Unfortunately, what counts as the "strong doctrine" is never made clear: when first introduced, it is "the doctrine, roughly, that all persons intrinsically 'desire' only power" (53). A few pages later we are told that one objection to the strong reading is that Nietzsche ultimately rejects WP 1067, which claims that "this world is the will to power—and nothing besides!" (57). This is a non sequitur: even if Nietzsche rejects this metaphysical claim about the ultimate constituents of reality, he could still accept a psychological claim about what people intrinsically desire. On the next page, the strong doctrine has changed yet again: it is now the view that the "feeling of power was the exclusive explanation for all human behavior" (58). So, over the course of five pages, we have at least three different things that are identified with the "strong doctrine of will to power":

(a) The only thing we desire for its own sake is power.
(b) The ultimate metaphysical constituent is power.
(c) The feeling of power is the exclusive explanation for human behavior.

These are different. You could accept (a) while denying (b) and (c). You could accept (c) while denying (a) and (b). And so on.

So one problem is that Leiter's target keeps shifting. But a deeper problem is that all of (a)-(c) are irrelevant in the dialectical context that Leiter has set. Remember, what is supposed to be under discussion is whether Nietzsche argues that power has a privileged normative status. Suppose Leiter is right that Nietzsche rejects (a)-(c). Why would that matter? Nietzsche could reject (a)-(c) and still hold that power has a privileged normative status. On my reading, for example, Nietzsche rejects (a)-(c) but nonetheless argues that power has a privileged normative status.

In sum: Leiter's defense of the anti-privilege reading is completely unconvincing. His targets are too amorphous and in any event do not bear directly on the question that he is trying to address. If we want to show that Nietzsche does not really believe the claims that he explicitly makes, we will need much more powerful arguments than this.

In fairness, I should note that the footnotes to this chapter do contain some engagement with more recent literature. But many of the notes on recent literature are either disappointingly imprecise or factually inaccurate. For example, in footnote 5, Leiter claims that my reading of Nietzsche on power fails because I "trade extensively on conflating" (i) whether Nietzsche employs power as an evaluative standard and (ii) whether Nietzsche thinks employing power as an evaluative standard is justifiable (51). If Leiter were a more careful reader, he would see that pages 153-6 of Katsafanas 2013 explicitly discuss just this distinction: as I put it there, "on the one hand, Nietzsche might think we can justify the claim that power has a privileged normative status. On the other hand, Nietzsche might give power a privileged normative status while maintaining that this stance cannot be justified" (Katsafanas 2013: 154). This is the very distinction that Leiter presents me as somehow failing to grasp.[4] And this misrepresentation is characteristic; I have mentioned another example above (falsely claiming that I treat will to power as the only human motivation), but I will give a third and fourth here. On pages 60-61 Leiter chastises me for not citing a book by Bernoulli, complaining that this book "is unreferenced in Katsafanas' discussion" (61); in fact I discuss Bernoulli's book on the very page that Leiter is complaining about (Katsafanas 2013: 248, fn. 8). And then there is this claim about Sharon Street:

> "unlike some recent Anglophone philosophers (e.g., Street 2006: 130–1), Nietzsche recognizes that evolutionary forces affecting human cognition do not necessarily prefer true to false belief—even in the case of ordinary knowledge about the empirical world" (87)

As someone familiar with Street's work, this claim struck me as surprising. So I checked. And sure enough, in the very article and on the very page that Leiter cites, Street makes the *opposite* claim from the one that Leiter attributes to her. She writes: "Sure, one might think, an organism who is aware of the truth in a given area, whether evaluative or otherwise, will do better than one who isn't. *But this line of thought falls apart upon closer examination*" (Street 2006: 130, emphasis added). She goes on to

---

[4] Perhaps Leiter thinks that although I discuss the distinction at length, I somehow fail to keep track of it. If so, we would need more than bare assertion: we would need some indication of where my argument is supposed to go astray.

write that that are many "kinds of truths such that it will confer either no advantage or even a disadvantage for a given kind of creature to be able to grasp them" (Street 2006: 130). So, just as I discuss the very distinction that Leiter accuses me of failing to notice and cite the very book Leiter accuses me of not citing, so too Street's paper provides a defense of the very claim that Leiter accuses her of failing to recognize. These are just a few examples chosen almost at random, but they are far from alone. Taken together, they do not inspire confidence that Leiter is an accurate reader of scholarship with which he believes he disagrees. Moreover, if Leiter gets the most basic factual claims about these views wrong, then in arguing against the views he will be tilting at windmills. And that is a shame, for Leiter is capable of clear, cogent arguments. So this feels like a series of missed opportunities.

I have been critiquing some of the crucial arguments in Leiter's book; but let me close with a final and more general remark. An overriding goal in Leiter's book is to make Nietzsche's claims align with those that Leiter considers respectable for contemporary philosophers (p. 9-11); indeed, he characterizes himself as aiming to "articulate and defend moral psychology in a Nietzschean spirit" (14). Leiter thinks that claims about power or the causal efficacy of conscious states are indefensible, for example, so he wants to read Nietzsche in a way that makes Nietzsche agree.

I do respect the intent: like Leiter, I think the most interesting historical work is not purely antiquarian, but addresses philosophical problems in their own right. But there is a delicate balance between showing that a historical philosopher has valuable contributions to contemporary debates and showing that the philosopher's views are *identical* in substance to those currently defended. I think the value of Nietzsche lies in his distinctiveness, rather than his alleged conformity to contemporary trends in social psychology and empirically oriented philosophy. After all, if (as Leiter argues) Nietzsche turns out to be someone who states in complex ways the same conclusions as contemporary philosophers and psychologists, why would anyone bother to read Nietzsche? Why spend years or decades unpacking Nietzsche's texts if all you end up with are the same claims and arguments that you could have gotten in an afternoon's perusal of some recent journal articles or popularized books on social psychology? To my mind, what rewards scholarship on Nietzsche is that he has *novel* views, that he reveals blind spots and deficiencies in contemporary work.

But this is very far from Leiter's approach. Throughout the book, Leiter's strategy is to slot Nietzsche into familiar categories: so, we have "the Humean Nietzsche" (6), the Calliclean Nietzsche (63), the Millian Nietzsche (53), "the Therapeutic Nietzsche" (7), the Wegnerian Nietzsche (141), the Haidtian/sentimentalist Nietzsche (81), and so on. While some of these categorizations are useful, I often felt myself wondering: what happened to the *Nietzschean* Nietzsche? If Nietzsche really were just a pastiche of past and future philosophers, why read him at all? To my mind, Leiter's book does not answer this question.

~ ~ ~

Mark Alfano's *Nietzsche's Moral Psychology* is both excellent and distinctive. The prose is clear and engaging; the exegetical work is exemplary; and many of the analyses are convincing. He employs a novel methodology in analyzing Nietzsche. And the new methodology really does open up some unexpected terrain: before reading Alfano, I found virtue-theoretical readings of Nietzsche unconvincing; now, I find them more tempting.

Let us start with the methodology, which Alfano presents in Chapter 2. It has four steps. First, we select "core constructs" for the study. These core constructs, Alfano says, should be chosen "by someone who is deeply familiar with the texts, has good intuitions about which constructs are associated in which ways, is aware of prominent and promising interpretations and suggestions already in the literature, and has the good fortune to be able to consult other scholars for leads" (14). The constructs that Alfano settles on include terms such as *type, drive, instinct, virtue, value, affect,* as well as particular instances of these more general categories, such as *courage, honesty, contempt, anger, disgust,* and so on. Alfano presents his list of core constructs on pp. 15-16; there are 48 such constructs.

The second step is to search the eKGW (via www.nietzschesource.org) for words that Nietzsche uses to refer to these constructs. So, for example, in order to find places where Nietzsche discusses the construct "honesty", Alfano searches for ehrlich* and redlich* (this returns all passages containing words that begin with "ehrlich" or "redlich"). Of course, in some passages Nietzsche does not use the German term to refer to Alfano's construct; more on this below.

Third, Alfano looks at the co-occurrence of the constructs within sections. How often does "honesty" appear with "drive"? How often does it appear alongside "courage"? (Let me register one worry about this methodology: as Alfano acknowledges, the methodology is sensitive to the size of the textual blocks that we choose. Rather than looking at how often constructs co-occur in a sentence or a paragraph, Alfano looks at how often they occur in a textual *section*. But the textual sections in UM are *much* longer than the textual sections in TI. So one worry is that many terms that co-occur in the lengthy sections of the UM will not co-occur in the short sections of TI. In principle, one could address this by treating smaller textual units [e.g., paragraphs, ten-sentence blocks, five-sentence blocks, etc.] as the level of analysis; it would be interesting to see whether this affects the results.)

Finally, Alfano visualizes the connections between key terms with semantic network maps. There are examples on page 25 and following. These maps show the "weighted degree" of the terms (the sum of all the co-occurrences of the construct), the degree of connection to other terms, and so on.

What are the results? It turns out that when Nietzsche is discussing moral psychology, the most prevalent concepts are, in order: *life, value, emotion, virtue, fear, conscience, instinct, courage, contempt,* and *laughter* (25). Some of this conforms to our antecedent expectations: I would have expected to find *life, value, instinct, emotion,* and so on. Others are more surprising: I would not have expected *laughter* to figure so prominently. Aside from mere frequency, the connectedness relations among these concepts are sometimes revealing: I would not have expected the strong link between *virtue* and *solitude,* or between *courage* and *health* (25).

I will mention two concerns about the methodology. First, in employing this methodology Alfano needs to decide which German words to treat as identifying the same concept and which to treat as identifying different concepts. Take *Trieb* and *Instinkt:* those are different words, but in my view they pick out the same concept. Alfano argues against this in Chapter Three, claiming that *Instinkt* refers to an innate drive, whereas *Trieb* refers to an acquired drive. Now, that is possible. I do not find Alfano's textual evidence for this point completely persuasive, but I can see room for disagreement here. But this raises a question: are the semantic network charts affected in significant ways by the decision to treat *Trieb* and *Instinkt* as distinct? That is, would we get importantly different conclusions if we were to treat these as one node rather than two? I do not know. But, if we would,

then it raises a worry about the methodology: if the semantic networks are substantially changed by contentious interpretive points, then the semantic network approach risks obscuring rather than resolving these points. That said, there would be an easy way to check: just run the semantic networking software twice, once with *Trieb/Instinkt* as one node and once with it as two nodes, and see what happens.

But things might grow more complicated with more complex concepts. Take *Leben:* that is one word, but in my view Nietzsche uses it to pick out several different concepts (sometimes he is just talking about living organisms; sometimes, about power; sometimes, about health; sometimes, about flourishing; sometimes, about expansive or procreative forces; and so on). If we look at the various semantic network visualizations in Chapter Two, *Leben* turns out to be the central node in the visualization of the corpus as a whole; moreover, it enjoys a central spot in the semantic networks for each of Nietzsche's published works. There is no work in which *Leben* is not one of the central constructs; and in many of the individual works, it is *the* central construct. So we might conclude that life is Nietzsche's most important notion in moral psychology. But what if *Leben* actually picks out—let us say—five distinct notions? If we could split up the occurrences of *Leben* into *Leben-1, Leben-2,* and so on, what would the network look like then?

Or, to press this problem from the opposite direction: what if Nietzsche sometimes uses two different words to pick out the *same* concept? Again, take *Leben.* Alfano claims that scholars including myself, Bernard Reginster, and John Richardson are wrong to give *will to power* such a central role in their interpretations: according to Alfano, will to power plays "only a modest role in [Nietzsche's] moral psychology" and has received "an undue amount of attention in the secondary literature" (26-27). Indeed, will to power is Alfano's primary example of a concept that has received too much attention in the scholarly literature (284). But I find this totally unconvincing. As Alfano acknowledges in a footnote on page 27, Reginster, Richardson, and I give power a central role in part because we take "will to power" and some occurrences of "life" to pick out the same notion. As I understand Alfano's methodology, he would rebut this charge by checking to see how often *Leben* co-occurs with "will*" and "zur Macht" (16). There is some co-occurrence, but perhaps Alfano takes the co-occurrence to be insufficient to warrant identifying these two terms. If that is the argument, I find it completely unpersuasive. If Nietzsche sometimes uses *Leben* to refer to will to power, then I would actually *predict* that we would find very few spots in which *Leben* and *Wille zur Macht* co-occur in the same textual section. Consider an analogy: Bentham equates utility with aggregate happiness; a textual analysis might (who knows) find that "utility" and "aggregate happiness" co-occur only a handful of times, specifically when Bentham is equating the two; but this miniscule degree of co-occurrence would not undermine, and could in fact support, the claim that Bentham equates the two terms. Just so with Nietzsche: "life" and "will to power" do not exhibit a high degree of co-occurrence because they are two terms for the same thing, so that using one term is sufficient. In short: if a philosopher uses terms T1 and T2 to refer to the same thing, we would not expect T1 and T2 to exhibit a high degree of co-occurrence. (Things are more complex with *Leben,* though: as I mentioned above, I think *Leben* picks out five or six different concepts, so we cannot simply assume that each occurrence of *Leben* can be replaced by *Wille zur Macht;* we have to engage in more sensitive, nuanced readings, holding that *Leben-1* is interchangeable with power, *Leben-2* is not, and so on.)

I have been focusing on particular constructs. But there is also a second, more general concern: we should not assume that *prevalence and interconnectedness* are correlated with *importance.* To simplify a bit, Alfano's digital humanities methodology looks at the frequency with which various concepts appear

and the degree to which they are connected to other concepts. There is a straightforward objection: a concept could be prevalent and interconnected but unimportant, and just so a concept could be infrequently mentioned but crucial. Here is a simple example: presumably the German word "und" would show up all over the place in Nietzsche's corpus and would be interconnected with all sorts of constructs, but clearly it is not an important philosophical term for him. Or, stepping back from Nietzsche: the term "categorical imperative" is not even close to being the most prevalent term in Kant's corpus, but it is undeniably one of the most important concepts in his practical philosophy. Or, if you do not like that example, consider an imaginary case: it is certainly coherent to imagine a philosopher who takes some concept C to be the single most important notion in her philosophical work, but who only mentions C a few times. (We could even test this with Nietzsche: he claims eternal recurrence is his most important thought, but I wonder how central it would appear in a semantic network chart.)

Of course, Alfano is aware of this. Claims about prevalence and interconnectedness are only the starting points: they give some evidence, but not *decisive* evidence, that a concept is important. So again this is a worry not so much about the methodology itself, as about the way in which the methodology can mislead: we should take these semantic network charts as starting points for reflection on Nietzsche's texts, rather than as decisively establishing any particular reading of the texts. This does suggest, though, that Alfano is overstating his case when he suggests that his digital humanities approach is the *best* approach to Nietzsche scholarship (13). I would say, instead, that while the digital approach sometimes can help resolve interpretive disputes, at other times it can be deployed effectively only after we have already resolved these disputes (e.g. with power and life).

I have focused on the methodology due to its distinctiveness. But, of course, the methodology is in the service of an interpretation. Briefly, Alfano argues that Nietzsche is a virtue theorist (Chapters 4-5); and then Chapters 6-10 offer analyses of several Nietzschean virtues (curiosity, courage, pathos of distance, sense of humor, and solitude). As Alfano summarizes it: Nietzsche is "an idiosyncratic philosopher of virtue who associates virtues less with rational dispositions acquired through habit and more with instinctual activity" (5). I lack the space to address the intriguing discussions of individual virtues, but I do want to examine Alfano's general notion of virtue.

Alfano claims that virtues are well-calibrated drives (93). More precisely, a drive counts as a virtue when it meets three conditions:

1. It is consistent with or supportive of life and health (93).
2. It does not "systematically or reliably end up inducing negative self-referential emotions" for aspects of the agent that are unmodifiable (94).
3. It does not "systematically or reliably provoke emotional condemnation from the agent's community" for aspects of the agent that are unmodifiable (94).

One potential shortcoming of this analysis is that Alfano, by his own admission, leaves *life* and *health* unanalyzed (97). He suggests that "we can start to make some sense of these constructs [life and health] by thinking in terms of full development, articulation, and expression of an agent's drives" (97). But I wonder about this. Nietzsche suggests that full development of certain drives hinder life or health. (Consider what happens when the ascetic develops his aggressive drives.) Nietzsche also suggests that we need to take account of relations between drives: certain configurations of drives will undermine life rather than promote it. Taken together, these points suggest that we do in fact need analyses of life, health, unity, and so forth.

Aside from that, I wonder which self-referential emotions qualify as *negative* (condition 2). One tempting thought is that an emotion counts as negative if it conflict with or undermines life and health; but this would make (2) largely redundant, as most of the work would be done by (1).

I think we can also question (3). Suppose a drive toward intellectual inquiry meets conditions (1) and (2) but fails (3): in particular, it systematically provokes emotional condemnation from a community of fundamentalist Christians. Surely it should still count as a virtue? More generally, why should we think that Nietzsche wants to relativize virtues to particular communities? Alfano does address this question. First, on 103-4, he argues that individuals who meet conditions (1) and (2) but fail (3)—but who, in different social circumstances, would meet (3)—fall under the Nietzschean type "criminal": for Nietzsche, "the criminal belongs to a type that, in a different context, might flourish as a hero. But the criminal finds himself in a situation where expressing his drives is almost certain to be disastrous" (104). And he elaborates on these points in his fascinating Chapter Five, which discuss the way in which drive-expression depends on "the semantics and valence of the labels applied to them" by social communities (116). There, he argues that condition (3) is important in part because someone disposed toward an end E will "end up thinking, feeling, and acting very differently depending on" how that E is understood by his community (116). While I appreciated Alfano's subtle discussions of the mechanisms by which this occurs, I was still tempted to reject (3) by saying that we can distinguish between what a community *calls* a virtue and what actually *is* a virtue. That is: why not allow that fundamentalist Christians will condemn an unfettered intellectual drive, agree with Alfano that this will produce damaging effects on the person who exhibits it, agree as well that the fundamentalist community will label it a vice rather than a virtue, but nonetheless say that it remains a Nietzschean virtue?

I have raised certain questions about Alfano's methodology and critiqued some aspects of his account of Nietzschean virtue. My largest concern is that if Alfano's interpretation is to give a full account of Nietzschean virtue, it needs to be situated within an account of life, power, and health. But a major contribution of this book is that it puts us in a position to address these questions, questions which we might otherwise have overlooked.

~~~

Aaron Ridley's *The Deed is Everything* focuses on Nietzsche's account of action. Ridley believes that traditional accounts of action make four claims: (1) that will and action are distinct; (2) that will causes action; (3) that actions are distinguished from other events merely in terms of their causes; (4) that will is something "inner", something that can be specified solely in terms of mental states and processes (p. 12). Ridley calls accounts that endorse these four claims *empiricist* accounts of action.

Ridley takes Nietzsche to endorse an alternative understanding of action, namely the *expressivist* theory. An expressivist theory of action holds: (5) that will and action are mutually constitutive; (6) that action must be understood teleologically; (7) that will cannot be specified without reference either to action, to a context of social norms, or to both; (8) "in acting the agent discovers something in or about himself (what he *is*, perhaps; or what he *really* willed)" (17).

Ridley takes (5) to involve a denial of (1) and (2), (6) to be a denial of (2), and (7) to be a denial of (4). If this is right, then expressivism would be an alternative to standard theories of action.

To understand expressivism's appeal, it helps to consider an example. Ridley gives several: trying to find the right words for an idea that you want to express (20-21), realizing in the course of conversation that you want your interlocutor to think well of you (38-39), creating various types of art (41-43 and 100-106), long-term challenging endeavors (124-131), and loving relationships (Chapter 5). Let's focus on a case of artistic production: Melville's writing *Moby Dick*. Here how an account embracing claims (1)-(4) would explain the action: (1') Melville engages in an episode of willing with the content *that I write Moby Dick,* which is distinct from but (2') causes his action of writing the novel; (3') his action of writing Moby Dick is distinguished from other events (such as inadvertently dropping his pen) merely in terms of its causes; and (4') this intention to write *Moby Dick* can be understood solely in terms of mental states and processes. This is a bad way of explaining the action because it is deeply implausible that intending to write Moby Dick consists in having some singular, separable mental state that stands in a causal relationship to the years-long project of writing the novel. A better view would be something like this: Melville intends to write a novel but only gradually figures out what this novel is; in so doing, he takes himself to gradually uncover what it is that he really intends; but then the intention cannot be specified in advance of the action, nor can it be specified independently of the action. And this might tempt us to embrace something like theses (5)-(8), above. After all, (5') it seems difficult or impossible to specify the content of Melville's willing without referring to the action that he is performing; (6') the action is specified in terms of its ultimate end, the production of *Moby Dick;* (7') we would not be able to specify what Melville wills without referring to what he is doing and the social context in which he is doing it; and (8') although Melville couldn't have specified in advance exactly what he was planning to do, he can see his intention realized in his completed novel. Of course, all of these claims would need further defense and elaboration, which Ridley provides in Chapter 1; but I hope this brief example suffices to get the picture into view.

Version of the expressivist theory have been attributed to Hegel and other philosophers. Robert Pippin (2011) and Alexander Nehamas (2018) have tried to read Nietzsche as endorsing some version of expressivism. But Ridley's version of expressivism is more modest: whereas Pippin and Nehamas suggest that expressivism is true of *all* action, Ridley wants to say "there are at least *some* actions that simply cannot be understood" according to the empiricist account, and so there is "reason to suppose that there are at least some actions that positively require" the expressivist account (39-40). So Ridley is restricting the thesis: the empiricist account is fine for some actions, but for others we need the expressivist account. This is a welcome modification, for it is hard to believe that simple actions, such as walking across the room to get a glass of water, require an expressivist account. However, I did not get a clear sense of what Ridley felt the ratio of empiricist actions to expressivist actions was: is Ridley claiming that *most* actions require an expressivist account? Or that *a few actions* do? It is not clear. Ridley does say that the types of action requiring the expressivist account "are really quite commonplace" (40); but again, this is rather vague. What are these *types* of action that require expressivist accounts? How frequently do they occur?

These questions might seem trifling, but I think they are important. A central topic in the early, Davidson- and Anscombe-inspired literature on action theory was the individuation of actions. Suppose I walk to my office. That takes around a half hour; along the way, I pass various cafes and shops, I see familiar faces, I might gaze at the trees or the sky, I might attend to the sound of the traffic. Is *walking to my office* one action? Twenty? A thousand? Does each footstep count as an action? Does each gaze at the sky qualify as a distinct action? For most purposes, the answers to these questions will not matter. But if we want to assess Ridley's claim that expressivist actions are

quite commonplace, we need to know how we are individuating actions. Here is the problem: if we individuate actions very finely (taking a step, moving one's head, etc.), it is possible that *no* actions will require expressivist analyses; whereas if we individuate actions very broadly (doing philosophy, writing a book, living a life), then it is possible that *all* actions will require expressivist analyses.

These points about action-individuation generate a problem for proponents of the expressivist view. Take a case of action that might seem to call for expressivist analysis, such as writing a book. The author does not know in advance what exactly he is going to write, how all the details will shape up, and so forth. But he does know that he wants to write a book. Ridley is correct in claiming that until the book is actually written we will not be able to specify the precise nature of the action. But at each moment, we will be able to specify the precise nature of the smaller actions that jointly compose the larger action of writing the book. So, the author does not know what is going to happen in the last chapter, but he does know that right now he is writing the introduction, etc. We can give simple empiricist accounts of each of these more immediate actions. So the proponent of the empiricist model can argue that the expressivist is focusing on the wrong things: the expressivist is trying to explain large scale, complex events; but he should, instead, focus on the smaller actions that compose these larger events.

More generally, it often seems to me that the examples that are supposed to motivate the expressivist picture are better analyzed in another way. If I understand correctly, the expressivist focuses on actions such as *writing a novel* and wants to say that, in some such cases, the nature of the intention cannot be understood until the action is complete. In one way this seems true: when the author sets out to write a book, she will not have every detail mapped out, she will often change her mind, revise things, perhaps scrap much of it and restart in a different way, etc. Eventually, if all goes well, she will end up with a book that seems to her a satisfactory or even a good one. But why should we say, with the expressivist, that her intention was to write this book that she has ended up with? I am tempted to say something else: that she had a general intention to write a good book, had thousands (or more?) of more localized subsidiary intentions, tried various courses of action, abandoned some, and eventually ended up with something that satisfied her. In other words, I suspect that what the expressivist is attempting to capture with talk of *intention* is really just some notion of *being satisfied with what you have done.*

In fact, we could phrase this as a philosophical worry about expressivism in general, rather than just Ridley's version of it: expressivism risks collapsing the distinction between *intending to X* and *being satisfied with X*. But those are distinct: there are many actions that I intended but with which I am unsatisfied, and there are many events with which I am satisfied but which were not intentional actions. Ridley does consider a version of this objection on pages 47-48, but his response is unsatisfying: he allows that agents can be mistaken about what they have done, and hence can confuse the claim *I intended X* with the claim *I am satisfied by X*. But Ridley takes this to be a specific instance of a more general point: first-personal avowals can be wrong. As he puts it, "although first-personal avowals do carry a certain prima facie weight in the context of expressive action, as they do everywhere, they are no more incorrigible there than they are anywhere else" (47). This does not address the real worry, though. The worry is that the expressivist has no means of distinguishing the following two cases: (a) I am satisfied by my X-ing, but my X-ing was not intentional; and (b) I am satisfied by my X-ing, and my X-ing was intentional. Or, looking at it the other way around: (c) I am dissatisfied by my X-ing, but my X-ing was intentional; and (d) I am dissatisfied by my X-ing, and my X-ing was intentional.

As I hope these remarks indicate, large-scale, temporally extended actions look difficult for the standard empiricist account only if we make two assumptions: (i) that the empiricist wants to account for long-term actions at this level of specificity; and (ii) that the agent is ultimately satisfied with his project because it expresses some intention that he had all along. I think both assumptions are false. With regard to (i), the empiricist can just say that he has no desire to explain actions such as *Melville's writing Moby Dick*. What he wants to explain, instead, is actions such as *Melville's deciding to write a novel,* or *Melville's drafting the first paragraph,* etc. That is, he wants to explain the parts of this more expansive event. With regard to (ii), it seems to me implausible to see the product—the novel *Moby Dick*—as what the agent intended all along. He intended something more amorphous, performed various actions in an attempt to achieve it, and eventually achieved something with which he was satisfied. There are puzzles concerning how the various smaller actions compose the larger event, how the intention to write *this sentence* relates to the intention to *compose the novel,* and so on; but I do not see how invoking the notion of expressive action helps with any of these puzzles.

I have been critiquing the account of expressive action, which occupies chapters 1-2. Chapters 3-4 extend this to an account of freedom: Ridley argues that certain commentators have gone astray in taking Nietzsche's discussions of freedom to center on claims about determinism; instead, Ridley offers some insightful reflections on the way in which Nietzsche might take artistic agency to be an exemplary expression of free agency. As Ridley puts it, Nietzsche's focus on artistry reveals that free action is

> an action in which my will is crystallized precisely *in* its realization. Therefore—since my action simply is, exclusively and without remainder, the expression *of* my will—there is no room, at this level, for my willing that the action were otherwise. The action is not only the action that I intend, but, in performing it, I discover exactly what my intention is. (105).

So for Ridley, Nietzschean free action is action that the agent does not and cannot will to be otherwise. Freedom is a kind of self-satisfaction or self-affirmation. This seems to me insightful: I think Ridley is correct that a notion of self-satisfaction plays a crucial role in Nietzsche's account of agency, though I read Nietzsche as linking this notion to unity rather than to freedom (see Katsafanas 2016, Chapter 7).

Ridley reads Nietzschean autonomy somewhat differently: to be autonomous "is to have the capacity to commit oneself to undertakings whose success-conditions are (at least partly) internal to the execution, and hence in coming to discover the determinate nature of one's will (i.e. to acquire self-knowledge)" (133). Here, I was somewhat less convinced: even if we grant that Nietzsche associates autonomy with internal success conditions and self-discovery, I do not see why we cannot account for these things with standard empiricist accounts of action. Similar questions arise in Chapters 5 and 6, which attempt to tie expressivism to Nietzsche's claims about morality and selfhood. That said, Ridley does make a convincing case for the claim that exemplary manifestations of will to power require performing the sorts of actions that look like good candidates for expressive analysis. Moreover, Chapter 5 contains nuanced and original discussions of the connection between a particular conception of love, on the one hand, and power, on the other. These sections are insightful and reward close study.

In sum: Ridley's book is intriguing, but I was unpersuaded by its central claims about expressive action. The book points to some complex agential phenomena; claims that these complex phenomena should be understood in terms of expressive accounts of agency; but never makes it entirely clear why we should think these complex agential phenomena are best understood as single

actions requiring expressive accounts, rather than as agglomerates of simpler actions that can be understood in the usual way. Nor does the book make it clear how the expressive account would illuminate (rather than simply use new terminology to describe) these complex agential phenomena. That said, this is the most detailed and developed analysis of expressive theories of action in the Nietzsche literature, and is a substantial improvement on earlier readings. It also contains valuable reflections on freedom, autonomy, and artistic agency.

~ ~ ~

I now want to stand back from the books and reflect on some more general concerns: how we should approach the task of interpreting Nietzsche; what we should make of Nietzsche's late views on ethics; and how moral psychology bears on Nietzsche's philosophical project.

Let us start with the point about interpreting Nietzsche. Each of these texts presents itself as rescuing Nietzsche from misinterpretation. Thus, Leiter wants to prevent Nietzsche from being 'moralized' (i.e., read in a way that makes him sound like a contemporary moral philosopher); Stern wants to prevent Nietzsche from being iron-manned (i.e., read in a way that assumes his arguments must be invulnerable to critique); Alfano wants to correct what he sees as a tendency to misrepresent Nietzsche's central concerns; and Ridley thinks writers have been misled when thinking about Nietzsche on action.

There is value in each of these points. For example, I do agree with Leiter that a great deal of modern moral philosophy is strikingly superficial, all too ready to accept uncritically the currently dominant ethos, all too reliant on intuitions. Nietzsche would see most of modern moral philosophy as laughable, as a mere apology for dominant values rather than a genuinely critical approach to them. But to my mind Leiter takes this too far: we should be wary of letting our animus toward contemporary moral philosophy prevent us from recognizing the way in which Nietzsche is developing a distinctive ethical view of his own, rather than merely braying "yea-yuh" to certain ideals. Nietzsche's concern, from *The Birth of Tragedy* to his last works, is not just to critique contemporary ethical views but to point toward an alternative. The alternative takes different forms as Nietzsche's views develop, but in the works from *Zarathustra* onwards it becomes increasingly clear that *will to power* and *life* are playing a dominant role.

Stern's view seems to me more accurate on this point: in the works from 1886 onwards, it is obvious that Nietzsche wants to use facts about power or life as evaluative standards. If you look at the late texts, it is just *very* hard to deny that this is so. Of course, this leaves open the question of whether Nietzsche's attempt to forge an ethical theory centering on life or power is successful (or even coherent). Stern claims that it is not; and Leiter agrees that any such project would be unsuccessful, which is part of what leads Leiter to deny that Nietzsche is really engaged in this project. I think this is a mistake.

To see why it is a mistake, consider Leiter's points about anti-realism. We should be cautious about the form of anti-realism that we attribute to Nietzsche. Leiter sees Nietzsche as embracing a relativist view according to which there are no ultimate justifications for preferring one set of evaluative claims to another (in this, Leiter agrees with post-modernist readings of Nietzsche, which he otherwise opposes). I have argued that Nietzsche's actual view is different: he accepts a form of anti-realism coupled with the claim that power or life is the ethical standard that we should employ. According to Nietzsche, we should embrace values that are life-enhancing or life-affirming rather

than values that are life-reducing or life-negating; and we are making a grave mistake when we fail to do so. (This is most explicit in passages such as *Antichrist* 2, but really the idea simply pervades the later works; Stern does an excellent job of highlighting this fact, and Alfano's textual analyses of *Leben's* role demonstrate the same point.) And I have claimed that Nietzsche actually does have good arguments in favor of that view: put simply, Nietzsche attempts to show that facts about our nature as drive-motivated creatures entail that we either embrace life-affirming values or end up in a conflicted and contradictory state (in which we at once strive for power and act in ways that undermine that very striving). This argument makes certain assumptions: it relies on a particular understanding of drives, power, life, and agency; it relies, too, on the idea that we have reason to prioritize unconflicted states over conflicted ones. But in my view it is a very powerful argument. Rather than ignoring it or prematurely concluding that it is unsuccessful, we should strive to understand it (see Katsafanas 2013, C. 6-8 for my reconstruction of it).

Which brings me to a final point. If I am right, Nietzsche's chief concerns – the way in which modern morality undermines power, the way in which modern culture imperils flourishing, the way in which we might strive for something more, the way in which this striving requires revaluation of the currently dominant values – are anchored in his moral-psychological claims. Increasingly, as his work progresses, Nietzsche relies on substantive claims about the nature of drives, motivation, consciousness, agency, and selfhood. He is interested in the critical role that these concepts can play, of course; he is keen to point out the way in which an adequate moral psychology will vitiate Kantian theses about agency, for example. But he also gives these concepts a positive role: he is eager to understand psychology in an unprejudiced, non-moralized way because he thinks that doing so reveals an alternative to the life-negating and nihilistic evaluative schemes that we have been operating under. In understanding ourselves, in understanding our psychologies, we (or, more accurately, some very small subset of us) can move toward values that enable the affirmation of life; and, more than that, we can hope that this process might lead toward new sets of values spreading and becoming culturally instantiated, thereby shifting all of us away from the decadent, life-negating, nihilistic culture that Nietzsche sees emerging. We stave off nihilism and life-negation through revaluation; revaluation, if it is not to falter or lead us astray, demands an understanding of moral psychology. In that respect, psychology is indeed the path to the fundamental problems.

These seem to me to be among Nietzsche's most original and sophisticated areas of thought. How should we analyze the concepts that play central roles in Nietzsche's ethical thinking (power, life, health, affirmation, negation, nihilism)? How should we understand the central topics in his moral psychology (drives, the conscious/unconscious distinction, selfhood, affects, and so on)? And, most of all: how do the central topics in his ethical thinking relate to the central topics in his moral psychology? For example, how does *will to power* relate to *drive?* How does *health* relate to *selfhood?* While the books above begin to probe aspects of these topic, and while I have offered a study of them my in own books (Katsafanas 2013 and 2016), these are rich grounds that merit further analysis.[5]

References

Katsafanas, Paul (2013), *Agency and the Foundations of Ethics: Nietzschean Constitutivism*. Oxford: Oxford University Press.

Katsafanas, Paul (2016), *The Nietzschean Self: Agency, Moral Psychology, and the Unconscious*. Oxford: Oxford University Press.

Katsafanas, Paul (2018), "The Antichrist as a Guide to Nietzsche's Mature Ethical Theory," in *The Nietzschean Mind*, ed. Paul Katsafanas, pp. 83-101. London: Routledge, 2018.

Nehamas, Alexander (2018), "Nietzsche, Intention, Action," *European Journal of Philosophy* 26(2): 685-701.

Pippin, Robert (2011), *Nietzsche, Psychology, and First Philosophy*. Chicago: University of Chicago Press.

Stern, Tom (2016), "'Some Third Thing': Nietzsche's Words and the Principle of Charity," *Journal of Nietzsche Studies* 47(2): 287-302.

Stern, Tom (2018), "Must We Choose Between the Real Nietzsche and Good Philosophy," *Journal of Nietzsche Studies* 49(2): 277-283.

Street, Sharon (2006), "A Darwinian Dilemma for Realist Theories of Value," *Philosophical Studies* 127(1): 109-166.