Nietzsche's Account of Self-Conscious Agency
Paul Katsafanas

Forthcoming in Constantine Sandis (ed.),
Philosophy of Action from 1500 to the Present Day, Oxford University Press

Draft of March 9, 2016

Philosophers discussing the role of self-consciousness in human action tend to fall into two camps. One side argues that self-consciousness gives rise to a rarified form of action. Locke, for example, claims that self-consciousness provides the agent with "a power to *suspend* the execution and satisfaction of any of its desires" (Locke 1689/1975, 263). Creatures who lack self-consciousness are simply buffeted about by whatever desire or instinct happens to predominate; self-conscious creatures, by contrast, can pause, examine their desires, and decide which one to act upon. As Locke continues, the self-conscious agent can "consider the objects of [these desires]; examine them on all sides and weigh them with others. In this lies the liberty that man has" (Locke 1689/1975, 263). So self-consciousness, in enabling a reflective suspension of desire and generating a capacity for choice, provides us with a distinctive form of liberty or freedom. Kant agrees, telling us that the self-conscious will "can indeed be *affected* but not *determined* by impulses... *Freedom* of choice is this independence from being determined by sensible impulses" (*Metaphysics of Morals* 6:213-214).

Another side downplays the importance of self-conscious thought, arguing that it plays a subsidiary role in human action. There are versions of this position in Hobbes, who identifies the agent's will with the motive immediately preceding action;[1] in Schopenhauer, who claims that while various motives may occupy our conscious thought, we have a "fixed disposition and unalterable character" that determines the way in which we will act on these motives (*On the Basis of Morality*, 112);[2] and in Hume, who famously argues that reason—including the operations of self-conscious thought—can have no original motivational impact on our actions.

Simplifying considerably, then, one side argues that self-consciousness drives human action and engenders a distinctive form of freedom; another side argues that we are actuated by passion or desire and that our actions are not different in kind from those of the brutes.

---

[1] When in the mind of man appetites and aversions, hopes and fears, concerning one and the same thing, arise alternately; and diverse good and evil consequences of the doing or omitting the thing propounded come successively into our thoughts; so that sometimes we have an appetite to it, sometimes an aversion from it; sometimes hope to be able to do it, sometimes despair, or fear to attempt it; the whole sum of desires, aversions, hopes and fears, continued till the thing be either done, or thought impossible, is that we call deliberation … In deliberation, the last appetite, or aversion, immediately adhering to the action, or to the omission thereof, is what we call the will… (Hobbes, Leviathan VI).

[2] In other words, character and motives jointly determine action; self-conscious willing adds nothing.This is complicated by Schopenhauer's distinction between empirical and intelligible character. Above, I'm focusing on empirical character. Schopenhauer has a notoriously obscure argument for the claim that while our empirical character is fixed, our intelligible character—our "will as the thing in itself"—is in fact chosen in an atemporal act of willing.

Many readers assume that Nietzsche enters into this debate, accepts its terms, and aligns himself with the second camp. It's not hard to see why. After all, Nietzsche's texts repeatedly deny that acts of will are the primary determinants of what we do:

> The error of false causality… We believe that we are the cause of our own will…Nor did we doubt that all the antecedents of our willing, its causes, could be found within our own consciousness or in our personal 'motives'… But today… we no longer believe any of this is true. The 'inner world' is full of phantoms and illusions: the will is one of them. The will no longer moves anything, hence does not explain anything—it merely accompanies events; it can even be absent. (TI VI.3)

We believe that our wills are causally efficacious, but our actions are actually caused by background forces that we fail to detect. Nietzsche sometimes goes further, claiming that that conscious thought, and in particular acts of will, are causally determined:

> Perhaps there exists neither will nor purposes, and we have only imagined them. The iron hands of necessity which shake the dice box of chance play their game for an infinite length of time; so that there *have* to be throws which exactly resemble purposiveness and rationality of every degree. *Perhaps* our actions of will and purpose are nothing but such throws… (D 130)

In addition, he claims that we are ignorant of most of our actions:

> however far a man may go in self-knowledge, nothing however can be more incomplete than his image of the totality of drives which constitute his being. he can scarcely name even the cruder ones: their number and strength, their ebb and flood, their play and counterplay among one another, and above all the laws of their nutriment remain wholly unknown to him. (D 119)

Given these claims about the background forces determining our actions and the ubiquity of self-ignorance, there's a temptation to slot Nietzsche into the second category of action theorists. Couple Hobbes or Hume with the idea that we're ignorant of much of what we do, and you have Nietzsche.

In this essay, I'll suggest that things are more complex. Nietzsche is actually questioning some of the tacit philosophical commitments that lead us into this apparent dichotomy. He thinks *both* sides are mistaken in important respects. Motives, acts of will, self-conscious thoughts, aspects of one's character, physiological features, environmental factors all play a role in constituting a nexus of forces that determine what we do. Self-conscious thought has a minor, albeit important and distinctive, role in this nexus. Its operations aren't what we expect: they are incremental and aggregative rather than sudden and decisive. In light of this, Nietzsche asks us to envision action in a strikingly different way.

In this essay, I'll attempt to get Nietzsche's theory of action into view, highlighting its central features and mentioning the ways in which it departs from standard accounts. Section One discusses Nietzsche's account of the opacity of human action. I focus on the the way in which the agent's experience of the world is shaped by unnoticed and unconscious factors. Section Two asks what role consciousness has in action. Section Three turns to the way in which Nietzsche

understands the action/behavior distinction. Finally, Section Four analyzes Nietzsche's account of freedom. What emerges is a view that is not just one more entry into the standard debates, but an attempt at rethinking the terms in which the debate is cast.

1. The opacity of action and deliberation

Perhaps the most striking feature of Nietzsche's account of agency is his emphasis on self-ignorance. His texts are simply overflowing with claims about our introspective fallibility, our dim awareness of our own motives, and our ignorance of our own deliberations. He claims that it is a "universal madness" to think that we generally know what we are doing. For "the opposite is precisely the naked reality demonstrated daily and hourly from time immemorial! … actions are *never* what they appear to be … all actions are essentially unknown" (D 116). Not only do actions contain hidden facets; we're often confused about the very distinction between acting and being acted upon:

> 'I have no idea what I am *doing*! I have no idea what I *ought to do*!' – You are right, but be sure of this: *you are being done!* [du wirst gethan!] at every moment! Mankind has in all ages confused the active and the passive: it is their everlasting grammatical blunder. (D 120)

This emphasis on self-ignorance is a pervasive feature of Nietzsche's texts. It invites a question: why does Nietzsche think that pointing out the various ways in which we're ignorant of ourselves has any philosophical relevance? After all, it's a truism that human action is obscure. Consider the range of philosophers who have embraced that idea: it's present in philosophers as disparate as Augustine and La Rochefoucauld, Montaigne and Spinoza, Rousseau and Schopenhauer. Even Kant, whose model of agency Nietzsche wanted to attack, emphasized that we can *never* be certain which motives we are acting upon (*Groundwork* 4:407).

Given that Nietzsche emphasizes the novelty of his remarks on self-ignorance, he must want to go beyond these familiar points. Below, I'll argue that the value of Nietzsche's account lies not in the mere assertion that human action is obscure, but in the particular account of how and why this is so. There are four important ideas, here: the existence of an unnoticed type of psychological state called the drive; the way in which drives impact conscious experience; the way in which experience is structured by the concepts the we employ; and, in light of these points, an argument for abandoning the typical focus on discrete causes of action. I'll treat these ideas in turn, below.

1.1 The importance of drives

Nietzsche introduces a type of motivational state that he calls the drive (*Trieb, Instinkt*).[3] He argues that drives are omnipresent in human action and complicate the structure of action. To see how, we'll need to consider four distinguishing features of drives.

First, while most motivational states take a one-part complement, drives take a two-part complement. Desires typically have one-part complements, in the sense that they motivate the agent to achieve some end (I desire a drink of water in order to relieve my thirst). Drives, on the other hand, admit a distinction between their aims (the processes of activity which the drive motivates)

[3] For an overview of the philosophical history of this concept, see Katsafanas (forthcoming).

and their objects (the things upon which these processes of activity are directed). For example, the aim of the aggressive drive is aggressive activity; its object might be some particular person.

Second, the drive is not sated by the attainment of its object; what it seeks is simply expression of its aim. The objects are adventitious, merely serving as opportunities for aim-expression. In other words, the aggressive drive differs from an ordinary desire in that it seeks continuous expression rather than the attainment of some state of affairs. The desire for water typically dissipates once the agent's thirst is quenched; the aggressive drive does not dissipate (and, Nietzsche suggests, in fact tends to increase in intensity) when it is expressed.

Third, drives operate by influencing the agent's perception and reflective thought. Drives generate affectively charged, selective orientations toward the world, which often incline the agent to see particular types of activity as warranted. We're familiar with the way in which desires and emotions can impact conscious experience: when I'm hungry, my attention is drawn to food; when I'm angry, I'm likely to find provocations in occasions that would otherwise be experienced as neutral. Drives are analogous. Suppose a drive is active, and seeks expression. If an appropriate object is unavailable, the drive will seek expression on whatever object happens to be present. The aggressive drive would most naturally be expressed upon things worthy of aggression. But, if there are no such objects, the drive will lead the agent to *seek* objects. It will do so by fostering configurations of affects and desires that lead otherwise neutral stimuli to be interpreted as worthy of aggression. For example, the drive may incline the agent to see the cashier's distraction as a personal snub, and thus worthy of a rude remark.

Finally, as the above example indicates, drives can arise independently of external stimuli. Once they have become active, they will seek discharge. The fact that drives are active and do not always arise in response to external stimuli makes their operations somewhat opaque to agents. When in the grip of a drive, I will experience the world as warranting certain kinds of response (aggressive responses, in the above example). But I may not notice that my experience of the world as warranting these responses is largely a result of the drive, rather than the features of the world itself. I'll say more on this below.

In short, then, the Nietzschean aggressive drive is a disposition that induces a configuration of affects inclining the agent to engage in aggressive activity; the agent in the grip of this drive is motivated not to achieve anything in particular, but simply to express aggressive activity (the drive's aim); this aggression will be directed toward someone or something (the drive's object); and the drive's arousal may be unrelated to external stimuli.[4]

Suppose we accept all of this. Then we can see that drives give rise to two pervasive forms of self-ignorance. First, they structure our thought such that we perceive certain ends as to-be-pursed. We confabulate, inventing reasons for pursuing the drive's aim, investing adventitious objects with the aura of significance. As Nietzsche puts it, a drive "erupts from time to time as reason and passion of mind; it is then surrounded by a resplendent retinue of reasons and tries with all its might to make us forget that fundamentally it is drive, instinct, stupidity, lack of reasons" (GS 1). So, for example, I think I'm responding rudely to the cashier because she's snubbed me; I don't see that the very perception of her as snubbing me is induced by a drive. Second, our actions don't have the

[4] Katsafanas (2016) for a defense of this reading.

motivational structure that we expect. We think we seek determinate ends and will be satisfied by their attainment; instead, we seek only processes of drive-expression and find attainment of ends at best temporarily satisfying. We're not just confused about the *content* of our motivation, but also about its *structure*.

1.2 Drives and experience

When searching for the causes of human action, we tend to focus on the decisions and intentions that precede the action. We notice these; and we notice, as well, any strong, forceful motives (lust, anger, love, hate, pity, fear, etc.) that figure in the etiology. Different philosophers weight these factors differently: the Kantian tradition focuses on the agent's maxims, the Humean on the agent's passions. Yet Nietzsche thinks that both camps miss the real determinants of our actions. For Nietzsche claims that the perspective through which we view the world plays a decisive role.

One way of understanding this point is by focusing on how drives operate. I'll use an example from Schopenhauer, who influenced Nietzsche's account of drives. Discussing the sex drive, Schopenhauer explains that the drive operates by occluding its aim:

> the sexual impulse, though in itself a subjective need, knows how to assume very skillfully the mask of objective admiration, and thus to deceive consciousness; for nature requires this stratagem in order to attain her ends. (WWR II: 535)

In other words, the sex drive *disguises* its true aim. That aim, according to Schopenhauer, is reproduction (WWR II: 535), but agents motivated by the sex drive are typically unconcerned with or opposed to reproduction. Thus,

> nature can attain her end only by implanting in the individual a certain *delusion,* and by virtue of this, that which in truth is merely a good thing for the species [i.e., reproduction] seems to him to be a good thing for himself, so that he serves the species, whereas he is under the delusion that he is serving himself. In this process a mere chimera, which vanishes immediately afterward, floats before him, and, as motive, takes the place of reality. This *delusion* is *instinct.* (WWR II: 538)

So the sex drive presents itself in a way that will incline the agent to act on it. Precisely because the individual does not have a conscious desire for reproduction (and may have a conscious desire *not* to reproduce), the drive presents its aim (sexual activity resulting in reproduction) in terms that will appeal to the agent. It makes potential sexual partners salient, it generates strong affects that incline us toward sexual activity, and so on.

The details of Schopenhauer's account might be disputed, but the only point on which I will rely is the central claim about *how drives operate.* As Schopenhauer puts it, "Here, then, as in the case of all instinct, truth assumes the form of a delusion, in order to act on the will" (WWR II: 540). In other words, the drive generates "delusions" in order to motivate the agent to attain its aim.

We might summarize these points as follows. The drive has some particular aim. The individual motivated by the drive is often ignorant of this aim—moreover, in some cases the individual would disavow the aim were he aware of it. Nonetheless, the agent pursues this aim. He pursues the aim

because the drive generates an affectively charged orientation toward the world that inclines the agent to experience pursuit of the drive's aim as appealing.

With that in mind, we might compare the attentiveness, the urgency, the desirous manner in which the sexually motivated individual sees his potential partners with the way in which an agent in a cool moment might perceive the same persons. What was alluring becomes indifferent; what was tempting is no longer even noticed; what was full of passion and desire recedes into the background.

The affectively charged orientation induced by the drive clearly has a tremendous impact on the agent's behavior; but Nietzsche thinks we tend to miss this. We think that the way we're viewing the world is determined by the world alone; what he wants to emphasize is that the way in which we experience the world is thoroughly altered by the operations of our drives.

These alterations as so profound that Nietzsche often compares the drive-induced-perspective/world relation to the sensory-stimuli/dream relation. *Daybreak* 119 offers an extended discussion of this phenomenon. Nietzsche starts with a discussion of dreams:

> Why was the dream of yesterday full of tenderness and tears, that of the day before yesterday humorous and exuberant, an earlier dream adventurous and involved in a continuous gloomy searching? Why do I in this dream enjoy indescribable joys of music, why do I in another soar and fly with the joy of an eagle up to distant mountain peaks? These inventions, which give scope and discharge to our drives to tenderness or humorousness or adventurousness or to our desire for music and mountains... are interpretations of nervous stimuli we receive while asleep, *very free,* very arbitrary interpretations of the motions of the blood and intestines, of the pressure of the arm and the bedclothes, of the sounds made by church bells, weatherclocks, night-revelers and other things of the kind. That this text, which is in general much the same on one night as on another, is commented upon in such varying ways, that the inventive reasoning faculty *imagines* today a *cause* for the nervous stimuli so very different from the cause it imagined yesterday, though the stimuli are the same: the explanation of this is that today's prompter of the reasoning faculty was different from yesterday's – a different *drive* wanted to gratify itself, to be active, to exercise itself, to refresh itself, to discharge itself... (D 119)

The sensory stimuli present from night to night remain relatively constant, while the dreams vary enormously. Nietzsche attributes the variation in dreams to the activities of different drives: the same sensory stimuli give rise to quite different dreams, depending upon which drives are most active. Having made this point, Nietzsche applies it to waking experience:

> Waking life does not have this *freedom* of interpretation possessed by the life of dreams, it is less inventive and unbridled—but do I have to add that when we are awake our drives likewise do nothing but interpret nervous stimuli and, according to their requirements, posit their 'causes'? that there is no *essential* difference between waking and dreaming? (D 119)

Nietzsche claims that just as drives influence the content of dreams, so too drives influence the content of waking experience. Of course, the connection between sensory stimuli and waking experience is *much* tighter than the connection between sensory stimuli and dreaming. But Nietzsche's point is that in waking, as in dreaming, our experiences are determined not by facts about the world alone, but also by facts about which drives are active. Thus, Nietzsche will speak of affects and drives as "coloring", "gilding", "lighting," and "staining" the world; these terms suggest that affects and drives highlight or even alter aspects of an experience, but not that they *create* the experience in the way that they create dreams (see for example GS 7, 139, 152, 301; BGE 186). Thus, Nietzsche is seeking to undermine the intuitively plausible thought that our perceptual experiences of the world are determined by nothing other than the nature of the world itself.

1.3 Perspectives

So drives induce affective orientations that alter the way in which we experience the world. Nietzsche extends this general point beyond drives, though, arguing that our conscious experience of the world is structured by the system of concepts that we possess and employ.

To see this, consider the conceptual shifts that Nietzsche describes in works such as the *On the Genealogy of Morality*. In that text, Nietzsche argues that ancient and modern moralities differ in their evaluations: what the ancients labeled good (strength, conquest, power, health, beauty, and so on) the early Judeo-Christian system labels evil; what the ancients labeled bad (commonness, ordinariness, humility, weakness, and so on) the early Judeo-Christian system labels good. Those points are familiar. However, Nietzsche also emphasizes that the ancient and modern moralities conceptualize agency, the self, freedom, and responsibility in strikingly different ways (see, for example, GM I.13). Though a full explanation of this point would take us too far afield, his claim is that agents who possess and employ different sets of concepts will experience the world in significantly different ways. Let me give one straightforward example.

Nietzsche claims that the Judeo-Christian moral system operates with a conception of value that valorizes expressions of weakness by interpreting them as expressions of strength:

> Weakness is being lied into something *meritorious*... impotence which doesn't retaliate is being turned into 'goodness'; timid baseness is being turned into 'humility'; submission to people one hates is being turned into 'obedience'... The inoffensiveness of the weakling, the very cowardice with which he is richly endowed, his standing-by-the-door, his inevitable position of having to wait, are all given good names such as 'patience', also known as *the* virtue; not-being-able-to-take-revenge is called not-wanting-to-take- revenge, it might even be forgiveness... (GM I.14)

In other words, actual manifestations of weakness are reinterpreted as valuable. The Judeo-Christian system attaches positive valuations to events that constitute reductions in power, and thereby induces agents to pursue reductions in power. If the agent accepts these conceptualizations, he can view his manifestations of actual weakness as chosen, and hence as expressions of power.

This is just one example of a pervasive phenomenon: the way in which we conceptualize or describe situations has a decisive impact on the decisions that we make, the things we find appealing, and the things we notice. Yet, Nietzsche thinks, we tend to overlook these effects. We focus not on the way the world is experienced, but on the decisions made in light of that experience. The experience is taken as fixed, when in fact it should be seen as mutable.

Thus, large swathes of Nietzsche's texts are devoted to tracing subtle social and culture influences on experience: witness his claim that our concepts of responsibility, agency, duty, morality, and so forth are influenced by now discredited religious assumptions (see the *Genealogy*), or his argument that particular valuations which come quite naturally to us (compassion as good, desire for power as bad, etc.) are conditioned by social practices. The claim, here, is that just as our drive-induced affects structure our experience of the world, so too do our concepts. And these affectively charged, conceptually laden experiences of the world largely determine what we do.

So Nietzsche's first claim about self-ignorance is, so to speak, about the optics of self-knowledge. Our attention is drawn to forceful, momentary passions, intentions, decisions, desires, and the like. Given their salience, these are the forces that we're inclined to treat as responsible for determining our actions. But, Nietzsche suggests, these experiences are actually downstream from conceptualizations of experience, from "perspectives." Once we've conceptualized the world in a certain way, most of the work of determining what we'll do is already over.

1.4 The artificiality of isolating motives

So far, I've emphasized several ways in which *experience* is taken for granted in traditional accounts of agency, but problematized by Nietzsche. The self-ignorance that Nietzsche focuses upon, then, is not the momentary lapses of introspection, not the inability to detect reasons or motives for particular actions. Rather, he's interested in the distortions that are introduced into our mental economy by our failure to appreciate the very structure of motivation.

Nietzsche deploys these claims in an effort to establish another troubling point: he argues that the quest to locate particular motives for action is largely misguided. For one thing, he argues that there is no definite set of motives for action. No matter how deeply we investigate, no matter how many motives we uncover for an action, we can always find a deeper layer. This is part of what Nietzsche means when he writes,

> Cause and effect: there is probably never such a duality; in truth a *continuum* faces us, from which we isolate a few pieces, just as we always perceive a movement only as isolated points, i.e. do not really see, but infer… (GS 112, emphasis added)

Here, Nietzsche suggests that we envision the specification of causes and effects—for our purposes, motives and actions—as artificial. We may specify some number of motives for an action, but there will always be more: there will always be different ways of understanding the causes, finer discriminations among the motives, different classifications or descriptions of them, and different prioritizations of them. Thus, Nietzsche writes,

> Everything which enters consciousness is the last link in a chain, a closure. It is just an illusion that one thought is the immediate cause of another thought. The events which are actually connected are played out below our consciousness: the series and sequences of feelings, thoughts, etc., that appear are symptoms of what actually happens! — Below every thought lies an affect. *Every thought*, every feeling, every will is *not* born of one particular drive but is a *total state*, a whole surface of the whole consciousness, and results from how the power of *all* the drives that constitute us is fixed at that moment — thus, the power of the drive that dominates just now as well as of the drives obeying or resisting. The next thought is a sign of how the total power situation has now shifted again (KSA 12: 1[61]/ WLN 60).

Our actions are products of our "total state" (*Gesamtzustand*), yet this total state cannot be adequately captured by talk of discrete motives or discrete causes and effects. For example, drawing on the discussions in the preceding sections, consider all of the factors that Nietzsche sees playing into the determination of action: conscious motives and intentions, to be sure, but also unnoticed mild affects, character traits, environmental factors, drives, the perspectives induced by these drives, the concepts with which we operate, and so on. We can single out one of these factors, focusing, for

example, only on the conscious intentions and motives, but these are simply a few of the factors impacting the action.

This is why Nietzsche is skeptical of the quest to find maxims, intentions, motives, and so forth, where these are conceived as discrete and self-sufficient causes of actions. To be sure, there are motives for action; but they acquire their significance only in light of the perspectives we adopt.

So we might, instead, picture action as brought about by a set of interacting forces, with each specified motive playing some role in determining the vector of the force, but none singly responsible for it. There's no *one* cause; there are a host of interacting causes that need to be understood as operating in tandem. Just so, Nietzsche claims, with human action. But our talk of *the* motive for action, *the* maxim, *the* intention, disguises this.

In all of these ways, then, Nietzsche goes beyond the humdrum assertion that our actions are opaque. Traditionally, the worry about self-ignorance is the concern that we might miss some of the motives for our action, or might misattribute our action to one motive (compassion, say) when it really springs from another (envy). Nietzsche isn't interested in this. He seeks to undermine our ordinary ways of conceptualizing human motivation.

2. The incremental nature of reflection's effects

The previous section emphasized the way in which our actions are opaque to consciousness. This raises a question about the status of conscious reflection. We often reflect upon our motives, the options lying open to us, the values we embrace, and make choices on the basis of this reflection. And we often end up acting in ways that we have chosen to act. Does Nietzsche's drive psychology and his account of self-ignorance commit him to denying this? Given his emphasis on the pervasiveness of confabulation, the superficial views that we have of our own actions, and our misunderstandings of what it is that we seek in action, it might seem so.

Some readers interpret Nietzsche as an epiphenomenalist about consciousness. For example, Brian Leiter writes, "*the conscious mental states* that precede the action and whose propositional contents would make them appear to be causally connected to the action are, in fact, epiphenomenal" (Leiter 2007, 10-11). And Mattia Riccardi agrees, attributing to Nietzsche the view that consciousness is "superfluous" because "a mental state has the causal powers it happens to have quite independently of its being or not being conscious" (Riccardi forthcoming, Section 3).

I've argued elsewhere that these readings are not defensible. There are at least three problems with interpreting Nietzsche as endorsing epiphenomenalism. First, the fact that we miss the complexity of our actions, coupled with the fact that non-conscious forces play a large role in determining our actions, does not even *suggest*—much less establish—that conscious thought has no important role in action. Think, in this vein, of the wealth of research showing that our beliefs are influenced by epistemically irrelevant factors such our moods, clutter on our desks, smells, and so on. This doesn't entail that our conscious deliberations about what to believe have no impact on our beliefs. Just so with deliberation about action: the fact that conscious phenomena are not the sole causes of our actions does not entail or even suggest that conscious phenomena play no causal role. As

Nietzsche puts it, even if our beliefs about our own motives and actions are highly inaccurate, they still have important effects on us. For example, in GS 44, Nietzsche writes:

> Important as it may be to know the motives from which humanity has acted so far, it might be even more essential to know the *belief* that people had in this or that motive, i.e. what humanity has imagined and told itself to be the real lever of its conduct so far. For people's inner happiness and misery has come to them depending on their belief in this or that motive—*not* through their actual motives. The latter are of second-order interest.

Nietzsche is offering a corrective to the accounts of human action that seem, to him, overly reliant on the import of conscious thought; but there is no indication that he's going to the extreme opposite position of denying that conscious thought has *any* role in action.

A second problem with the epiphenomenalist reading is that it saddles Nietzsche with a sharp distinction between the conscious and the unconscious, with the conscious having special properties (such as causal inertness) not shared by the unconscious. This has no real grounding in his texts; Nietzsche emphasizes the continuity between the conscious and the unconscious. For a simple example of this, consider the discussion of drives, above: Nietzsche treats drives as unconscious forces with conscious manifestations. The drive acquires its efficacy in part by fostering a particular conscious configuration of experience.

Third, Nietzsche's texts constantly appeal to conscious thoughts having important causal effects. I won't review the textual evidence here, but to give a sample, consider the following passage, which emphasizes the way in which conscious interpretations alter their objects:

> *what things are called* is unspeakably more important than what they are. The reputation, name, and appearance, the worth, the usual weight and measure of a thing—originally almost always something mistaken and arbitrary, thrown over things like a dress…has, through the belief in it and its growth from generation to generation, slowly grown onto and into the thing and has become its very body: what started as appearance in the end nearly always becomes essence and functions [*wirkt*] as essence! … Let us not forget that in the long run it is enough to create new names and valuations and presumptions in order to create new 'things'. (GS 58; cf. GS 44)

For a concrete example of this, consider the *Genealogy*. There, Nietzsche argues that the ascetic priests offer religious views that constitute interpretations of suffering. He is concerned that these interpretations *alter* the motivational propensities of the affects. It follows that conscious thought is causally efficacious: the interpreting of our affects plays a causal role in the production of action. As Nietzsche puts it elsewhere, "that a violent stimulus is experienced as pleasure and pain is a matter of the *interpreting* intellect … and one and the same stimulus *can* be interpreted as pleasure or pain" (GS 127).

In light of these considerations, it's clear that Nietzsche is not an epiphenomenalist. Nonetheless, it's obvious that Nietzsche thinks traditional interpretations of consciousness are getting something wrong. How, then, should his remarks on conscious thought's connection to action be interpreted?

Although providing the textual evidence for this would necessitate a much longer essay, I've elsewhere argued that Nietzsche is responding to a model of agency associated with Kant and Locke. Suppose we see Kant and Locke as endorsing the following three claims:

> (Suspension) When an agent reflects on her motives for A-ing, she suspends the influence of the motives upon which she is reflecting.
>
> (Inclination) In deliberative agency, motives incline without necessitating. The agent's motives could be the same, and yet she could choose differently.
>
> (Choice) Typically, if I am faced with two actions that it is possible for me to perform, A-ing and B-ing, and I choose to A, then I will A.

Nietzsche rejects the Suspension claim. He claims that the agent's reflection is "secretly guided and channeled" by his drives and affects (BGE 3). As the previous section indicated, Nietzsche sees drives and other motivational states as manifesting themselves by coloring our view of the world, by generating perceptual saliences, by influencing our emotions and other attitudes, and by fostering attractions and aversions. Although reflection may appear to suspend the effects of motives, it typically fails to do so; the influence of the motives simply becomes more covert, operating through reflection itself.

Although Nietzsche rejects Suspension, he accepts Inclination and Choice. Given his emphasis on conscious thought's ability to shift the motivational direction of our drives and affects, it's clear that changes in the agent's reflective thought can lead to changes in what the agent does. In this sense, motives incline without necessitating: if we hold motives constant and vary conscious thought, we sometimes get different actions. So, too, with the point about Choice: Nietzsche doesn't deny that agents typically act in accordance with their choices; he simply points to the way in which these choices are themselves influenced and constrained by background factors that the agent typically fails to appreciate.

A reading of this form is not only more philosophically cogent, but fits better with Nietzsche's texts. He isn't concerned with the choice-action connection, but with the way that the conscious choices are driven by factors unknown to or unnoticed by the agent.

3. The action/behavior distinction

So Nietzsche does not deny the causal efficacy of conscious thought. But let's consider how this point relates to Nietzsche's account of willing. It's one thing to say that the will—the agent's capacity to engage in self-conscious episodes of choice—has some causal effects, and it's quite another to offer a determinate characterization of what these effects are. Are the effects significant enough that there will be any philosophically relevant difference between willed and unwilled actions?

It might seem not. On the Nietzschean model, actions are the product of a vector of forces that can include drives, affects, and conscious thoughts; when present, these conscious thoughts may play

only the smallest of roles in determining the nature of the action. Why, then, should it matter whether the conscious thoughts are present?

In fact, Nietzsche doesn't think that there is any philosophically significant difference between actions whose etiology includes an episode of conscious willing and those whose etiology does not. However, Nietzsche is interested in something that philosophers have (he thinks mistakenly) attempted to capture by speaking of conscious willing: the distinction between *genuine action* and *mere behavior*. Unlike many other philosophers, he does not align this distinction with the willed/unwilled distinction. In other words: genuine actions can be unwilled, and willed actions can be mere behavior.

A word on this. Some commentators have tried to read Nietzsche as rejecting the action/behavior distinction. However, Nietzsche is explicit about his reliance on such a distinction. Not only does he tell us that activity [Aktivität] is one of his foundational concepts [Grundbegriffe] (GM II.12), he also repeatedly relies upon a distinction between genuine actions and their degenerate relatives. He writes

> Nothing is rarer than a *personal* action. A class, a rank, a race, an environment, an accident—everything expresses itself sooner in a work or deed, than a 'person'. (KSA 12:10[59])

Some actions are products of agents or persons; others, the vast majority, are products of forces operating in or through the person. There are many examples of this claim in Nietzsche's texts; see, for example, GM II.12, TI VI.2, and TI VIII.6.

So Nietzsche accepts the genuine action/mere behavior distinction, but denies that it should be explained by appeal to acts of willing, reflection, or deliberation. Instead, Nietzsche employs a notion of *agential unity*. To cite just one example, he treats Goethe as exhibiting the requisite form of unity:

> What [Goethe] wanted was *totality*; he fought against the separation of reason, sensation, feeling, and will [*das Auseinander von Venunft, Sinnlichkeit, Gefühl, Wille*] (preached with the most abhorrent scholasticism by *Kant*, Goethe's antipode); he disciplined himself to wholeness... (TI IX.49)

Nietzsche claims that genuine actions are those that exhibit this form of unity, whereas mere behaviors are those that lack unity.

But how does Nietzsche analyze the notion of unity? A common way of interpreting Nietzschean unity is in terms of predominance: unity is achieved when some set of drives predominates and imposes a kind of order on the other drives. Thus, we might say that an agent with a coherent set of drives performs genuine actions, whereas an agent with conflicted, disorganized drives does not. Although this reading is standard in the literature, it seems to me mistaken.

There are two problems. First, there is no obvious reason for assuming that the dominant part of the self has some special claim to being expressive of the self. We often distinguish actions produced by the agent from acts caused by the agent's dominant motive. When the struggling alcoholic is overcome by his craving for alcohol, when this craving subordinates all of his other motives, we don't treat this as exemplary of genuine agency. Yet the predominance model would suggest that it is.

But there is also a second problem with the predominance model: there is textual evidence that Nietzsche dissociates unity and dominance. He derides those who become "as a whole the victim of some detail of us [*als Ganzes das Opfer irgend einer Einzelheit an uns werden*]" (BGE 41), and I've argued elsewhere that he presents individuals such as Wagner as disunified yet controlled by a dominant drive (see Katsafanas 2016).

What, then, is Nietzsche's account of the action/behavior distinction? Notice that in the quotation from KSA, above, Nietzsche emphasizes the rarity of *personal* actions. What occupies Nietzsche's attention is the degree to which the action is an expression of the agent herself, rather than forces operating through the agent. When Nietzsche suggests that a class or an environment expresses itself through the agent, he has in mind the sorts of scenarios that are ubiquitous in his works: the scientist takes herself to be dispassionately committed to the truth, but is additionally motivated by ascetic tendencies (GM III); the Christian takes himself to be committed to compassion, but is covertly exercising his desire for dominance (GM I); and so on. In short, agents are mistaken about the forces operating through them, and are often motivated by forces that, upon reflection, they'd disavow. It is not mere ignorance that matters, but ignorance that, once revealed, shifts the agent's attitude toward the action.

I've agued that we can best characterize this kind of unity in terms of the degree of conflict between the agent's reflective thought, on the one hand, and his drives, on the other. In particular:

> (Nietzschean Disunity) The agent A's, and affirms his A-ing. However, if the agent had further knowledge of the drives and affects that figure in A's etiology, the agent would not affirm A-ing.
>
> (Nietzschean Unity) The agent A's, and affirms his A-ing. Further knowledge of the drives and affects that figure in A's etiology would not undermine this affirmation of A-ing.

With these points in mind, we can see that disunity constitutes a form of psychic conflict. An agent acts, and approves of his action. However, this approval is contingent upon ignorance of the drives and affects that are actually leading him to act. So there is a conflict between the agent's attitude toward the action as he takes it to be, and the agent's attitude toward the action as it is. Moreover, disunity implies that one has affects and drives that are moving one in ways that one would disavow. Thus, there is an interesting form of conflict between the agent's reflective and unreflective aspects at the time of action.

To the extent that this unity condition is met, the agent counts as performing a genuine action. Notice that unity, so described, doesn't require extensive self-knowledge, nor does it require that the agent engage in episodes of reflection, deliberation, or explicit choice prior to action. In these

respects, it fits nicely with Nietzsche account of the way in which action is typically brought about (see Section One).

4. Freedom for the benighted

So far, we've seen that Nietzsche emphasizes the pervasive self-ignorance that are present in action; that he treats conscious thought as having an important, but not decisive, role in action; and that he analyzes the action/behavior distinction in terms of unity. Let's now ask how these accounts of the psychological constituents of the person come together in an account of the agent herself.

Nietzsche treats selfhood as an aspirational term: we are not selves merely in virtue of being human. Rather, Nietzsche claims that selfhood is something that must be attained. For example, UM III.1 distinguishes the "true self" from what the person already is. Elsewhere in that work, Nietzsche writes, "Be yourself! All you are now doing, thinking, desiring is not you yourself" (UM III.1). GS 335 claims that we "want to become those we are", and urges people to "create themselves." Indeed, Nietzsche claims that most of us *lack* selves: "we absolutely should not assume that many human beings are "people" [Personen]," he tells us (KSA 12:10[59]).

So what, exactly, is involved in the transition from lacking to having a self? Nietzsche typically associates genuine selfhood with a form of self-determination. The genuine self is not simply buffeted about by the values and customs of her society; instead, she critically engages with her values, rejecting some and accepting others. She manifests a form of freedom. Nietzsche regularly speaks of "evaluating on one's own," being "sovereign," and being "autonomous" (HH Preface 3, GM II.12). He writes that the free individual "is obliged to have recourse to his own law-giving" (BGE 262), and that free individuals enjoy a "constraint and perfection under a law [*Gesetz*] of their own" (GS 290; cf. GS 347).

So self-determination seems criterial for genuine selfhood. But what exactly does Nietzsche mean by self-determination? Self-determination or autonomy is typically analyzed as the capacity to govern one's actions according to principles or values that one has adopted for oneself, instead of principles or values that are externally imposed. We can see why Nietzsche would be attracted to such a view: he not only praises those who achieve freedom from conventional morality, but also urges us toward the "creation of our own new tables of what is good" (GS 335, italics removed). However, accounts of autonomy vary tremendously in their details. Kant, for example, claims that being autonomous requires acting on the categorical imperative; some maintain that genuine autonomy requires libertarian freedom (see, for example, Kane 1996); those influenced by Plato claim that autonomy requires government by reason rather than passion (see, for example, Watson 1975); and so on. Nietzsche emphatically rejects these conceptions of autonomy (see, for example, GM II.2). So it's not sufficient simply to note that Nietzsche treats genuine selfhood as requiring self-determination; we need to say more about what, exactly, self-determination would be for him.

Though a full treatment of this point would necessitate a paper of its own, we can make some preliminary points. Freedom, for Nietzsche, consists in being the source of one's own values. Nietzsche believes that human beings have acquired the capacity to regulate their actions via consciously adopted principles and goals. However, most human beings can only regulate themselves in this way by depending on external standards, customs, and sanctions. A human being

counts as free when she is able to regulate her action without dependence on these kinds of external props. This is the sense in which Nietzschean freedom is self-determination. As external influences are not transparent or obvious, genuine self-determination requires self-understanding. We must track down and analyze the ways in which external factors surreptitiously influence us. In short, we must cultivate self-understanding: "to be allowed to have a say about value and disvalue, one must see five hundred convictions *beneath* oneself, -- *behind* oneself" (A 54).

The task of freedom, then, is to become a source of value, out of self-understanding. Thus, Nietzsche envisions the agent attaining a nuanced, comprehensive knowledge of the influences upon her values, actions, and thoughts, and attaining a critical distance from them.

But how do we mark the external/internal distinction? In other words, suppose I attain comprehensive knowledge of the origins of some particular value. What next? What criteria do I draw on in assessing the value? There are different ways of interpreting Nietzsche on this point. Subjectivist readings treat him as having no real answer: the choice of which values to accept and which to reject will be ultimately arbitrary. Another reading, which I've argued enjoys more textual support and philosophical cogency, interprets Nietzsche as arguing that critical assessment is conducted in terms of a standard he calls "will to power". Thus, he writes that the "principle of revaluation" or the "standard by which the value of moral evaluation is to be determined" is "will to power" (KSA 12:2[131]), and in a section entitled "my conception of freedom," he claims that freedom is measured according to the degree of power expressed by an individual (TI IX.38).

Nietzsche doubts that most individuals will be capable of attaining freedom (GS 18, GS 98, GS 347), and he presents it as a dangerous aspiration for most of us (EH II.9). His account of motivation's pervasive effects on reflection entails that most attempts to assess values will be driven by unnoticed background affects and commitments; that, in short, these allegedly autonomous assessments will be a farce, driven by the very values the agent purports to suspend. Difficult as it may be to achieve, though, this is his ideal. The self-determining agent acquires deep, comprehensive knowledge of her values; she assesses these values in light of Nietzsche's criterion of will to power; and she acts on the basis of this assessment.

5. Conclusion

I've given a brief overview of Nietzsche's account of action. We've seen that Nietzsche's account centers on four points. First, he emphasizes the opacity of action, offering sustained analyses of the particular ways in which reflection is driven and distorted by non-conscious factors. Second, he seeks a theory of action that treats conscious reflection and deliberation as having important, but merely incremental, effects on the agent's actions. Third, he develops a novel account of the action/behavior distinction, severing that account from traditional reliance on claims about the import of conscious willing and deliberation. Finally, he argues that selfhood should be treated as an aspirational concept, realized when the agent attains a form of self-determination that requires both self-understanding and critical engagement with one's values.

References

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

Katsafanas, Paul (forthcoming), "The Emergence of the Drive Concept and the Collapse of the Animal/Human Divide," in Peter Adamson and G. Fay Edwards (eds.), *Oxford Philosophical Concepts: Animals.* Oxford: Oxford University Press.

Leiter, Brian (2007), "Nietzsche's Theory of the Will," *Philosophers Imprint* 7(7): 1-15.

Riccardi, Mattia (forthcoming) "Nietzsche on the Superficiality of Consciousness," in Manuel Dries (ed.), *Nietzsche on Consciousness and the Embodied Mind.* Berlin: Walter de Gruyter.